PROCEEDINGS OF THE 1970 CLINIC ON LIBRARY APPLICATIONS OF DATA PROCESSING

UNIVERSITY OF
ILLINOIS LIBRARY
URBANA CHAMPAIGN
LIB. SCI.

NOTICE: Return or renew all Library Materials! The *Minimum* Fee for each Lost Book is ~~$50.00~~ $125

DEC 1 8 2000

The person charging this material is responsible for its return to the library from which it was withdrawn on or before the **Latest Date** stamped below.

Theft, mutilation, and underlining of books are reasons for disciplinary action and may result in dismissal from the University.
To renew call Telephone Center, 333-8400

UNIVERSITY OF ILLINOIS LIBRARY AT URBANA-CHAMPAIGN

LIBRARY AND
INFORMATION SCIENCE
LIBRARY

L161—O-1096

PROCEEDINGS OF THE 1970 CLINIC ON LIBRARY APPLICATIONS OF DATA PROCESSING:

MARC USES AND USERS

Papers Presented at the
1970 Clinic on Library Applications
of Data Processing
Conducted by The University of Illinois
Graduate School of Library Science
April 26—29, 1970

PROCEEDINGS OF THE 1970 CLINIC ON LIBRARY APPLICATIONS OF DATA PROCESSING: MARC USES AND USERS

Edited by

KATHRYN LUTHER HENDERSON

University of Illinois
Graduate School of Library Science
Urbana, Illinois

COPYRIGHT 1971

The Board of Trustees of the University of Illinois

LC CARD NUMBER: 65-1841

ISNB: 0-87845-019-X

Proceedings of the Clinics since 1963 are available from the Illini Union Bookstore, 715 S. Wright Street, Champaign, Illinois 61820. The Proceedings for 1963 through 1967 are $2 paper and $3 cloth; 1968 and 1969 are $5.

C64p
1970
cop. 3

TABLE OF CONTENTS

FOREWORD: THE MEANING OF MARC vii
Carl M. Spaulding

INTRODUCTION ix
Kathryn Luther Henderson

THE EVOLVING MARC SYSTEM: THE CONCEPT OF A DATA UTILITY 1
Henriette D. Avram

MARC USERS: A STUDY OF THE DISTRIBUTION OF MARC TAPES AND THE SUBSCRIBERS TO MARC 27
Hillis Griffin

THE BNB MARC PROJECT 36
R. E. Coward

MARC BOOK CATALOG PRODUCTION IN WASHINGTON STATE 53
Josephine S. Pulsifer

MARC AT RICE UNIVERSITY 66
Frederick H. Ruecking, Jr.

MARC UTILIZATION IN THE UNIVERSITY OF CHICAGO LIBRARY BIBLIOGRAPHIC DATA PROCESSING SYSTEM .. 78
Robert S. McGee and Robert C. Miller

COMMERCIAL USERS OF MARC: AUTOCOMP, INC. 98
Fred C. Cole, Jr.

SUMMARY OF MARC 105
Frederick G. Kilgour

INDEX 111

FOREWORD: THE MEANING OF MARC

In recent years the cost of academic libraries has been increasing more rapidly than other costs of education. Because many functions performed in the operation of libraries are highly repetitive and/or clerical in nature the electronic computer has been an obvious candidate for the task of controlling the inflationary conditions. Many of the attempts to apply the computer to library problems did not recognize the size nor complexity of these problems and consequently failed or bogged down badly. One undertaking which has achieved significant results is the MARC development which began with a pilot project involving the Library of Congress and sixteen other libraries and evolved to a weekly service by LC to more than sixty subscribers to the MARC tapes.

MARC was conceived in consequence of an early appreciation by its planners and supporters of the centrality to library automation of machine-readable bibliographic records in a standard format. While the number of libraries now utilizing the MARC tapes on a regular basis is small and may remain so for a while, the system remains important because of its achievement in development of a standard format and because it, both directly and indirectly, is bringing about the accumulation of an increasingly large computer manipulable bibliographic data base. Without such a data base, opportunities to apply new technologies to bibliographic functions would in most cases have to be passed up for reasons of economics.

Here are a couple of examples of such technological application situations:

Outputting bibliographic information from computers has generally been done by means of high-speed line printers. There are a number of relatively permanent limitations of line printers including speed, character set, number and quality of copies, bulkiness of output, and cost. Fortunately just now an alternative to the line printer, computer output microfilming (COM), is experiencing very rapid development.

Even at this early stage, COM units record at rates of ten, twenty, and even thirty thousand lines per minute, which is one to two orders of magnitude faster than line printers. Since many COM units generate characters electronically there is the potential for nearly unlimited increase in character set size. Speed of operation and the small amount of film needed for microform publications insure low first copy cost, inexpensive use copies and permit frequent

republication from updated files. With newly introduced microfilm readers utilizing cassettes, it is possible to produce very compact and convenient book catalogs quickly and at low cost. Already several catalogs of this type are being produced and used, and others certainly will be in the next few years. Without a machine-readable data base this would not be possible.

At the University of Illinois a new type of low cost, on-line interactive terminal is being developed for Plato IV, a computer-based educational system. Plato IV is designed to provide varied education services from a large-scale, time-shared computer to as many as 4,000 students at one time. One of the applications being considered for the system is bibliographic searching. However, until catalog information is in computer-usable form, no bibliographic application will be possible.

Carl M. Spaulding
Program Officer
Council on Library Resources, Inc.

INTRODUCTION

In April 1969, the Library of Congress began to distribute to other libraries its cataloging services on machine-readable tape. This event has often been compared in importance to that of 1901 when the Library first began to distribute its cataloging services. Because of the importance and potential of this new service the eighth annual Clinic on Library Applications of Data Processing, conducted by the Division of University Extension and the Graduate School of Library Science of the University of Illinois, April 26-29, 1970, departed from its past consideration of all major aspects of library operations to concentrate on this specific aspect which can influence many different library operations. The intention was to review MARC after one year of operation, to present the current picture and future programs of the Library of Congress in regard to MARC and to assess the local, national and international potential of this service.

Carl M. Spaulding, program officer, Council on Library Resources, Inc., opened the conference with remarks about the meaning of MARC and its potential.

Henriette D. Avram was well qualified to speak on the purpose of a data-oriented, computer-based centralized service, with emphasis toward generalized applications on a centrally maintained set of data files for access by a variety of users. It is within such a data utility that the MARC system is "evolving" at LC.

Hillis Griffin writes from personal experience with the MARC distribution service and as a MARC user. He praises some MARC subscribers for their hard work with MARC but also chides others for lack of initiative in using MARC. He concludes that in this business "there is really no other road to success than through hard work."

The next series of papers consists of reports from some of the hard-working MARC subscribers referred to by Griffin. The British MARC program and its development, as well as a critical look at other aspects of MARC, are noted by R. E. Coward. Josephine Pulsifer puts MARC within the context of a number of cooperative efforts in Washington State.

Although illness prevented Frederick Ruecking, Jr., from attending the conference, his paper is included here. In it he explains the broadening of the original perspective for MARC in the Fondren Library, Rice University.

Robert S. McGee and Robert C. Miller report on MARC usage in

the University of Chicago Library's bibliographic data processing systems.

At the clinic two commercial users of MARC also reported: William R. Nugent, for Inforonics, Inc., and Fred. C. Cole, Jr., for Autocomp, Inc. Cole's paper is included in the published proceedings.

Frederick G. Kilgour "puts it all together" with his summary paper pointing out the successes of MARC, the problems with MARC and looks to its evolution in the future.

Herbert Goldhor, director, Graduate School of Library Science, helped with the planning of the conference. Robert D. Kozlow, automation librarian, University of Illinois, Hillis Griffin, and Frederick G. Kilgour also made valuable suggestions. Donna Duff Lenfest, clinic supervisor, helped in a variety of ways. All of their efforts are well deserving of acknowledgment for contributing to the success of the clinic.

Kathryn Luther Henderson
Clinic Chairman and Editor

Henriette D. Avram
Assistant Coordinator of Information Systems
Information Systems Office
Library of Congress

THE EVOLVING MARC SYSTEM: THE CONCEPT OF A DATA UTILITY

Many documents have been published describing the current MARC system, the procedures and computer programs that are used at the Library of Congress (LC) for the MARC Distribution Service and other projects for divisions within LC.[1] This paper attempts to explain the rationale of the system, to summarize what exists today (the results of the efforts of 1967-1969), and to describe the emerging plans for the next generation of the MARC system (1970-1971).*

THE CONCEPT OF A DATA UTILITY

To describe clearly what exists today and the plans for the future, it is necessary to explain the thinking in 1967 that led to the present MARC system. At that time, the only requirement for an operational system was the MARC Distribution Service. However, the designers of the system assumed (with nothing more concrete to base their assumptions on than intuition of what might be just around the corner) the use of one system for many purposes. The procedures and software necessary to input, format, validate, retrieve, update, and output cataloging records for the MARC Distribution Service are basically the same as those required to process all forms of material to produce bibliographic tools for use at LC.

The decision to design a generalized MARC system, while engaged in the analysis needed to functionally describe the man-machine specifications for the Central Bibliographic System (CBS),† forced us to try to understand the relationship of the MARC system to the total automation program of LC. We accepted the premise, and

---

*System as used in the remainder of this paper is concerned primarily with the computer hardware/software configuration rather than the procedures for editing, inputting, etc.

†A program consisting of seven phases. Phase I: Survey of the Present Manual System; Phase II: System Requirement Analysis; Phase III: Functional Description of Recommended System; Phase IV: System Specifications for Equipment and Software; Phase V: System Design; Phase VI: Implementation of New System; and Phase VII: Operations of a New System.

still do to date, that work accomplished would provide the prerequisite experience and knowledge for more advanced efforts. We fully recognized that the totality of the CBS and all it implies would evolve over many years, that the hardware/software configuration and the man-machine interfaces were not fully understood. The requirement to implement the MARC system in an environment of many variables and unknowns resulted in the conceptualization of a centralized data processing service called a data utility.* The hope was that, within the constraints imposed by the complexity of a many-faceted information system, our progress would be as orderly as possible and our efforts would have meaning for the future system.

The term "data utility" is used to describe a data-oriented, computer-based centralized service, with emphasis toward generalized applications on a centrally maintained set of data files for access by a variety of users. This concept differs significantly from a computer utility. A computer utility allows many users at remote sites to use a central computer concurrently. Each user operates as though the computer were dedicated to his problem alone. In general, each user produces his own programs and maintains his own data files. He may choose his own programming languages; he may be compiling a program, debugging a program, or executing a program; he may request output on any available device, e.g., printer output, teletype, cathode ray tube, or plotter. The data utility is similar in that certain services will be available at remote sites, but all take the form of interrogating existing files with existing software. The user does not produce the application programs or maintain the data files. (Certain files will be updated on-line by qualified library personnel, e.g., MARC editors will correct and verify records on-line.)

The operations of a library are file oriented. Any approach toward the automation of library processes must be based on the improved handling of files. These files must be converted to machine-readable form and maintained. It must be possible to access these files in a variety of ways both for maintenance and for retrieval of information, and to provide a variety of products for use by the staff and/or the public.

Within the data utility framework, files are maintained in a single standardized format, i.e., the records within each file are self-defining and the data elements in each record are represented in a common structure. The LC MARC processing format[2] was specifically developed for such a role, has already been used for several files, and, by the nature of its design, can be used for many others.

---

*The author wishes to acknowledge the work of Coyle & Stewart, Computer Applications Consultants, who, under contractual support, assisted the Information Systems Office in formalizing this concept, as well as some of the advanced planning described in the remainder of this paper.

The centralization of the computer facility and the standardization of the record format across all files have allowed development of general purpose application programs. These programs operate on the data without concern for the fact that there are a variety of files. Thus, it has been possible to serve different interested library personnel, the users of the data utility.

This generalized concept, although more difficult to implement, has paid off many times. The following briefly describes some of the LC projects that are either in operation or in the check-out phases that use the MARC system for the processing of the data from input through output.

1. A variety of book catalogs have been produced for internal use of the Geography and Map Division. The off-line retrieval subsystem has been used to extract records pertaining to a particular subject matter for specialized catalogs.

2. Files are maintained for the reference collection of the Science and Technology Division reading room. This collection contains current and non-current material made up of monographs, serials, and technical reports. Book catalogs are issued for use by the reference librarians.

3. A bibliography is compiled by the Science and Technology Division for the U.S. Army Cold Regions Research and Engineering Laboratory. Quarterly and annual reports are published as well as a monthly current awareness listing.

4. The National Directory for Latin Americanists, published by the Hispanic Foundation, is processed through the MARC system. Members of the foundation staff also use the retrieval subsystem to answer specific queries from the data base.

5. A file containing citations of journal articles, government documents, congressional documents, monographs, and internal Legislative Reference Service (LRS) reports is maintained. In addition to the production of book catalogs, this file supports a selective dissemination of information (SDI) system providing "current awareness" based on profiles for the staff members of LRS in their support of Congress.

A data utility encompasses both batch and on-line processing. Its advantage is that the service is open-ended in that new applications and data files may be added with facility whenever required. Although the MARC system was designed as a batch processing, tape-oriented system, the data utility concept represented a systematic framework for the expansion of its capabilities and services.

THE CURRENT MARC SYSTEM

The current MARC system is defined as four subsystems: input, file maintenance, retrieval, and output. The programs within each

subsystem are generalized, data independent where possible, table driven where the function of the program assumes known characteristics of the data (e.g., validation of MARC tags for a given form of material such as maps), and parameterized where the user, within constraints, can specify his requirements (e.g., format control for printing of bibliographic listings).

The system is a batch processing tape system operating under DOS on a system 360/40. Figure 1 is a high-level flow diagram, each rectangle representing a complete program.

Following the numbers associated with each rectangle in the diagram, the functions of the programs are briefly described below:

1. *Preedit.* Accepts source data from a variety of input devices (punched paper tape, magnetic tape, punched cards) and converts the data from the code of the input device to EBCDIC code for 360 processing. The preedit program deletes all function codes (codes from the input device to control the input device and/or to indicate to the computer program end of field, end of record, etc.), and writes each field of a MARC record on tape as a physical record. The present input process at LC consists of a magnetic tape inscriber through a converter to computer-compatible magnetic tape.

2. *Fmtedit* (*Format Edit*). Processes each record from preedit output (each record is a field of a logical MARC record) and produces a pseudo-MARC processing record. The format edit program eliminates all redundant fields (last field in is the one retained which allows the input typist to replace an entire field while typing) and corrects all characters, words or lines that the typist corrected by the use of backspace and delete codes during the typing operation. A field containing the order of input of the MARC record is prefixed to each format edit record. Both Preedit and Format Edit are data independent. They will accept and process any variation of the LC MARC processing format following the prescribed input procedures.

3. *Conedit* (*Content Edit*). Accepts the format edit output, determines the form of material (monograph, serial, map, etc.) being processed, activates the table of tags, indicators, subfield codes (hereafter referred to as content designators), and any special program modules required for the processing of that form of material. Content designators are validated and tags converted from a mnemonic to numeric value. (The editor has the option of using either form of tagging.)

Frequently appearing indicators and subfield codes do not require explicit editing or keying. The content edit program recognizes their absence and sets them automatically. Record and field level error conditions are flagged in the record, which is written on the content edit output tape and the processing results tape. All input records to Content Edit are written on the processing results tape, but depending on the type of error, they may or may not be written on the content

edit output tape. For example, if the LC card number is incorrect (discovered through recomputing the check digit), there is no way to correct the record in error since all control is by the card number. Therefore, the record is written on the processing results tape only.

4. *Sort.* Sorts the content edit output tape in LC card number sequence for further processing.

5. *Update1.* Accepts the sorted content edit output tape which may contain new, delete, replace, or verification records. The tape is processed against the MARC working file (old) which contains records that have not been declared error free. An updated MARC working file (new) is generated containing all new and modified records (resulting from the correction of fields) in addition to all records for which there was no transaction. All verified records and correction and deletion transactions which found no match on the working file (old) are transferred to the master file transaction tape for later processing. Transaction records which have an error in the transaction record itself, as well as all corrected records, are written on the processing results tape.

6. *Print Index.* Lists the LC card number and the date of the last transaction of records on the working file (new) for the MARC editorial office.*

7. *Update2.* Accepts the master file transaction tape which contains newly verified records and correction and deletion transactions which found no match on the working file (old). The tape is processed against the master file (old), and an updated master file (new) is generated containing all verified and modified records in addition to all records for which there was no transaction. All records modified on the master file, transaction records which have an error in the transaction record itself and all transaction records for which no match was found on the master file (old), are written on the processing results tape. Deletion records are processed against the master file (old) in two ways:

a. Deletion records referring to records not yet distributed to MARC subscribers cause the deletion of the record.

b. Deletion records referring to records that have already been distributed to MARC subscribers result in a change of the status of the original record. These records are transferred and remain on the master file until the next distribution cycle when they are transmitted to the subscribers and deleted from the file.

8. *Print Index.* Lists the LC card number and the date of last transaction of records on the master file (new) for the MARC editorial office.

9. *Sort.* Sorts records contained on the processing results tape

*The office in the LC Processing Department where MARC records are edited, keyed, corrected, and verified.

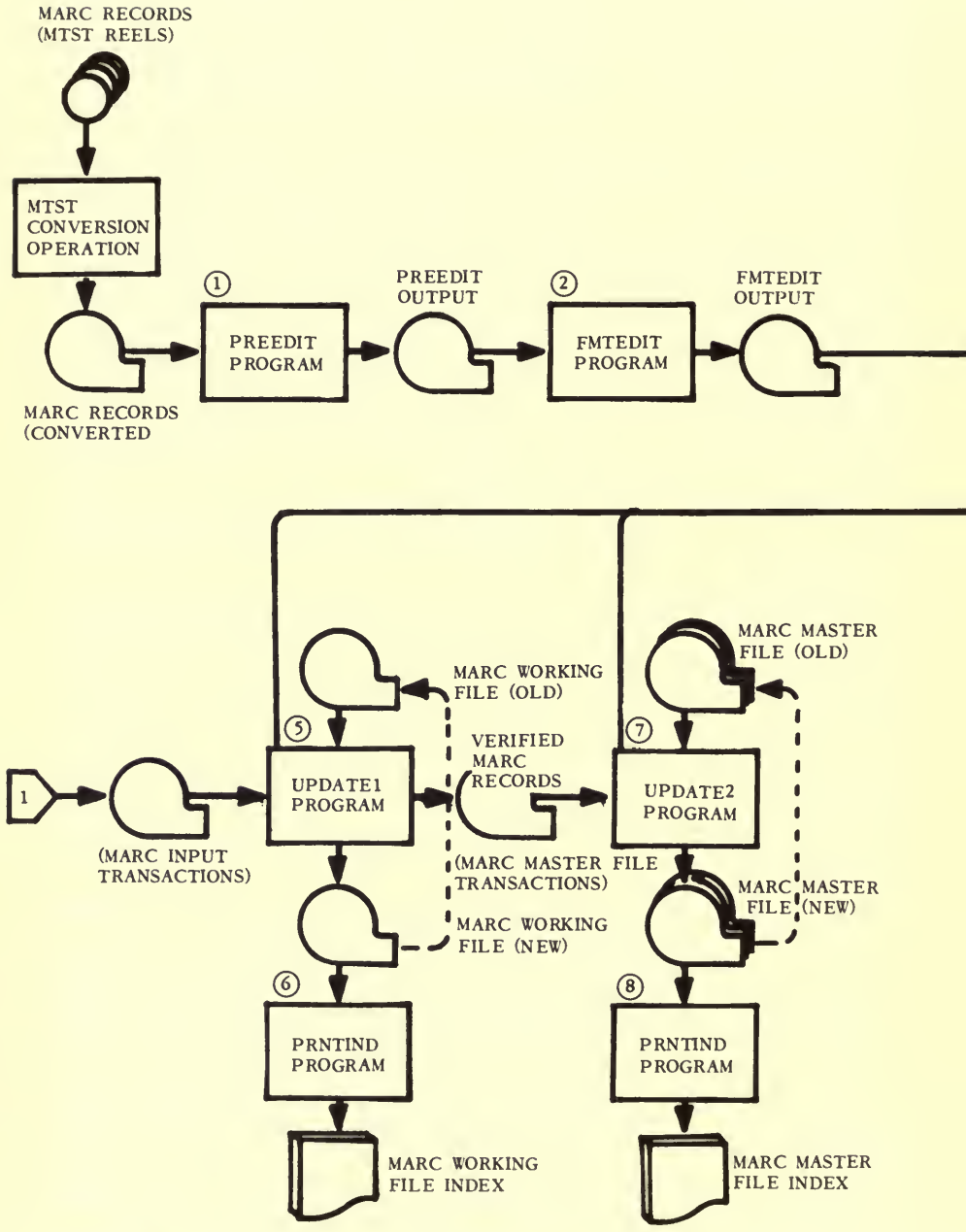
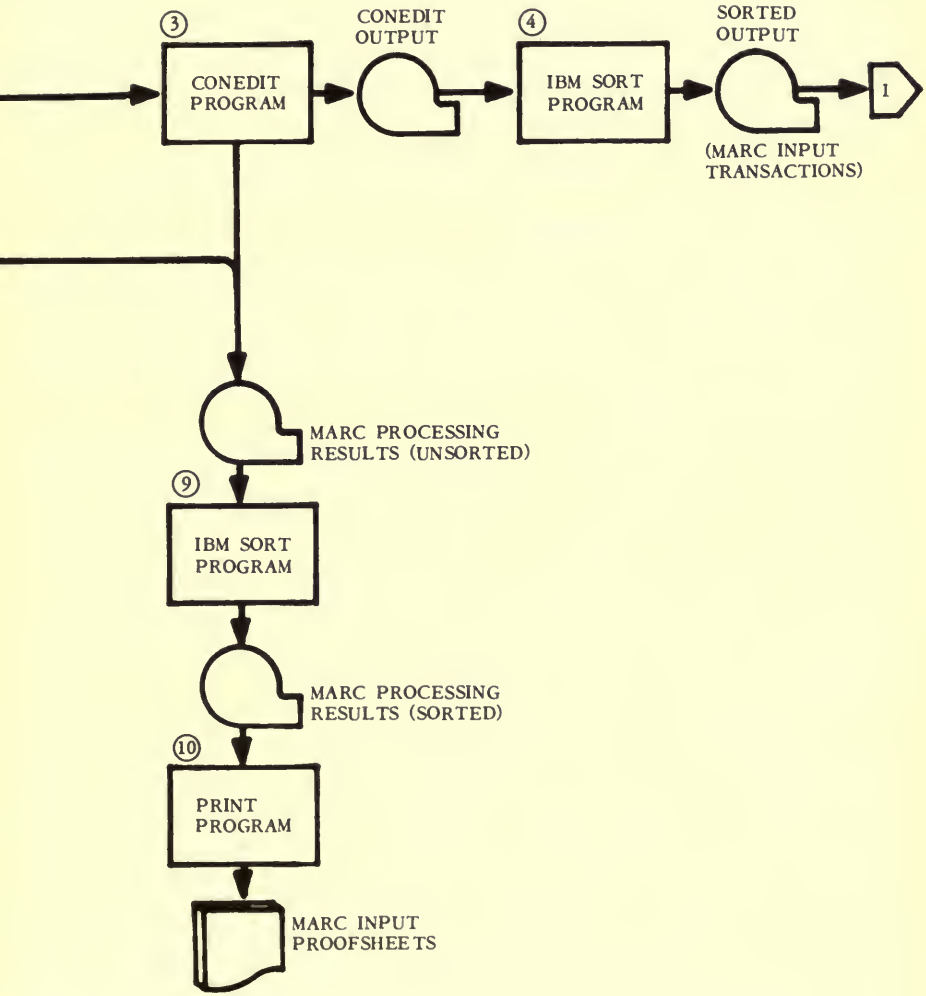


Figure 1. Current MARC II Processing System


3
CONEDIT PROGRAM
CONEDIT OUTPUT
4
IBM SORT PROGRAM
SORTED OUTPUT
1
(MARC INPUT TRANSACTIONS)
MARC PROCESSING RESULTS (UNSORTED)
9
IBM SORT PROGRAM
MARC PROCESSING RESULTS (SORTED)
10
PRINT PROGRAM
MARC INPUT PROOFSHEETS

(output of Content Edit, Update1 and Update2). The records are sorted on the field prefixed to the record during format edit processing containing the order of input. This sort serves two purposes: 1) the subsequent proofsheet run is arranged in input order, thus facilitating the matching of the proofing record with the original input worksheet; and 2) in the proofsheet run, the listing of each transaction record appears with the record it corrected. Thus, the entire transaction history of the record is displayed for analysis and for any subsequent action by editors.

10. *Print Program.* Displays all records contained on the sorted processing results tape. The contents of the record are displayed in an order determined to be the most useful for proofing against the original input worksheet. Record and field level error messages are generated to assist in the management of records within the system.

FURTHER DEVELOPMENT OF THE MARC SYSTEM

THE NEED FOR EXPANSION

The current MARC system is tape oriented in keeping with the hardware configuration available at LC at the time of implementation. LC's commitment for the operational MARC Distribution Service did not allow sufficient set-up time for the acquisition of additional equipment or the design of software required for the effective use of random access devices.

The present procedures for original input of data and correction of mistakes in editing or keying result in a cyclic process. The records are input, a proofsheet is provided, corrections are made on the proofsheet, and then they are rekeyed and run through the system producing another proofsheet. This procedure is continued until the record is declared error free. At that point, the LC card number is keyed into the system and the verified record is removed from the working tape and stored on the master tape. Control procedures are required to monitor the movement of the record from original input to the final cycle when it is considered verified.

The experience at LC indicates that on-line input of new records is not required but there is a definite need for on-line correction and verification of records on the working file and the master file. This will reduce the paper work involved, expedite the processing of records, and should result in an increase in the number of records processed. The processing time required to revise records in a tape system, the efficiency offered by on-line correction and verification, and the eventual requirement for more than one access point to a record are the factors that initiated the redesign of the MARC system. The redesign was strongly influenced by the desire to build upon what was already available and by the desire to provide a solid base for the expansion of the automation program.

FORMAT DESIGN

The MARC processing format will remain basically the same for the expanded system. Additional control codes will be added to facilitate increased demands.

Two approaches have been taken in specifying the MARC formats to fit within the concept of the data utility:

1. Care has been exercised to identify data common to different kinds of material with the same content designators. Unique content designators are assigned for data elements peculiar to any one form of material, e.g., the scale of a map for the single sheet map format.

Table 1 correlates the tags used in formats designed to date at LC. The monograph format is used in the present MARC Distribution Service; the serial format has been published to elicit comments; the map format will be published in the near future; and the motion picture and filmstrip format is still in the design stages. All tags have not been included in Table 1; only a sufficient number are listed to demonstrate how tags are assigned.

2. Specific codes have been assigned to the record for a variety of control purposes. For example, a code is used to indicate the currency of a record in relation to the record in the official catalog. This allows the initial input and processing of a large number of records which could include records requiring comparison with the official catalog. These records can subsequently be extracted from the machine-readable file for this purpose. In addition, this code provides a mechanism for controlling the distribution of records to MARC subscribers. A retrospective record converted by the RECON Pilot Project may be input and processed, but not distributed until the machine-readable record is compared with the record in the official catalog.

Other examples of the use of control codes are the source and distribution codes. The source indicates the originator of the machine-readable record (this does not in all instances mean the originator of the cataloging). For example, the source code "Main Reading Room" signals a record for a work in the Main Reading Room reference collection, although the cataloging was actually performed by the LC Processing Department. In contrast, the source code "Geography and Map" means both a record for a single sheet map in the Geography and Map Division and also that the Geography and Map Division did the cataloging. The destination code indicates the file that the record will reside on. In the case of the Main Reading Room collection, the records could reside on the MARC master file, the way records reside on a map master file. If the Main Reading Room records reside on the MARC master file, they actually become part of the future LC machine-readable official catalog; yet they can be extracted to produce a book catalog of the reference collection, recognizable by the source code.

Table 1. Tags for Representative Data Elements in Four Types of Bibliographic Records

TAG	NAME	Mono-graphs	Serials	Maps	Motion Pictures and Filmstrips
200	Title as it appears on piece		✓		
210	Abbreviated entry or title		✓		
240	Uniform title	✓	✓		
241	Romanized title	✓		✓	✓
242	Translated title	✓			
243	Collective title				
245	Title statement	✓	✓	✓	✓
246	Varying forms of title		✓		
247	Former titles		✓		
260	Imprint	✓	✓	✓	✓
261	Production and release				✓
265	Subscription address		✓		
400	Series note, personal name	✓	✓	✓	✓
410	Series note, corporate name	✓	✓	✓	✓
411	Series note, conference	✓	✓	✓	✓
440	Series note, title	✓	✓	✓	✓
490	Series, traced differently or untraced	✓	✓	✓	✓
500	General notes	✓	✓	✓	✓
501	Bound with notes	✓			✓
502	Dissertation notes	✓			
503	Bibliographic history note	✓		✓	
504	Bibliography note	✓		✓	
505	Contents note	✓		✓	✓
506	Limited use note	✓	✓		✓
507	Scale note			✓	
508	Credits note				✓
510	Indexing and abstracting coverage		✓		
511	Cast note				✓
515	Explanation of dates and volumes		✓		
520	Abstract or annotation	✓			✓
525	Supplement note		✓		
530	Additional physical forms available		✓		
555	Cumulative indexes		✓		
700	Added entry, personal name	✓	✓	✓	✓
710	Added entry, corporate name	✓	✓	✓	✓
711	Added entry, conference	✓	✓	✓	✓
730	Added entry, uniform title	✓	✓	✓	✓
740	Title traced differently	✓		✓	✓

An encoding level code has been defined and is currently in use to indicate if the machine-readable record is the result of assigning content designators with the book in hand or from a surrogate record. Current MARC is an example of assigning content designators with the book in hand, and RECON records, an example of assigning content designators from the LC printed card. Should LC proceed with pre-publication cataloging, another encoding level would be assigned to indicate this condition, i.e., the bibliographic description is not completed. In all three cases, current MARC records, RECON records, and pre-publication records, the source and destination would be "Processing Department" and "MARC master file" respectively.

Since the organization of the machine-readable records of LC is not known at this time, these control codes will provide information which may be needed for a specific choice of organization at a later date. In the interim, one system, namely the MARC system, can manipulate all the files.

The careful assignment of content designators and the use of the control codes insure compatibility for processing, take advantage of the utilization of the same program modules whenever possible, route the record to the proper file after processing through the same series of programs, and identify records in sufficient detail to facilitate their integration within a future system.

EXPANDED MARC SYSTEM

The expanded MARC system will be the composite of several subsystems. It will include what exists today (a batch tape subsystem), it will be modified to use disk instead of tapes as storage (a batch disk subsystem), and finally, it will include on-line capability for correction and verification. Any one of the subsystems may be used, depending on the file or records being manipulated.

Figure 1 remains intact as the batch tape subsystem of the composite system. Figure 2 is a schematic representation of the expanded system, and each rectangle represents a program. Section I will operate with section II as a subsystem (batch disk subsystem) and also with section III as another subsystem (on-line correction and verification subsystem). The input of new records always occurs in an off-line mode whether the file resides on tape or disk. The input of correction and verification records may occur through section I (off-line) or section III (on-line). Table 2 lists the subsystems involved and the mode of input. Access to records is by LC card number because of the immediate requirements for the MARC Distribution Service. However, it is anticipated that the on-line system will be expanded to 1) on-line input for other than the correction and

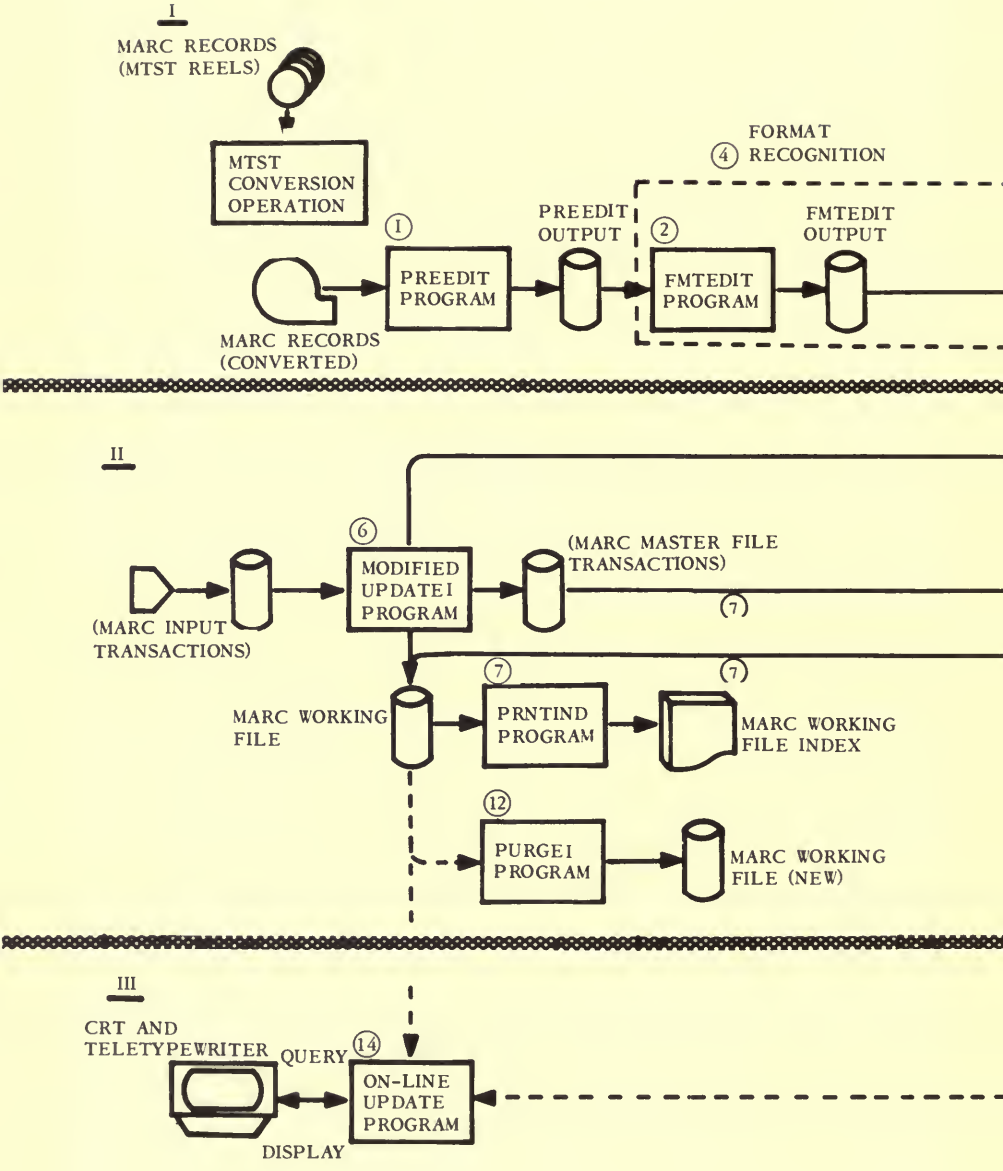


Figure 2. Expanded MARC Processing System

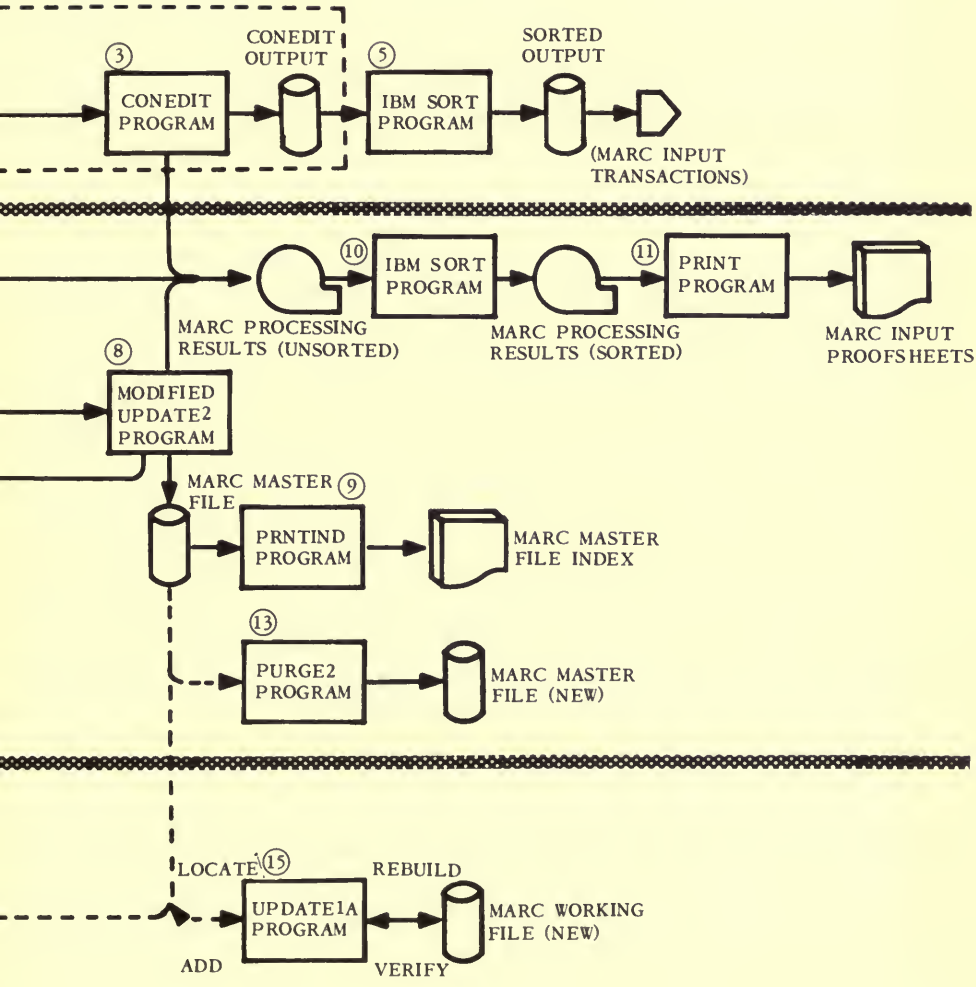

CONEDIT OUTPUT
3
CONEDIT PROGRAM
5
IBM SORT PROGRAM
SORTED OUTPUT
(MARC INPUT TRANSACTIONS)
10
IBM SORT PROGRAM
11
PRINT PROGRAM
MARC PROCESSING RESULTS (UNSORTED)
MARC PROCESSING RESULTS (SORTED)
MARC INPUT PROOFSHEETS
8
MODIFIED UPDATE2 PROGRAM
MARC MASTER FILE
9
PRNTIND PROGRAM
MARC MASTER FILE INDEX
13
PURGE2 PROGRAM
MARC MASTER FILE (NEW)
LOCATE
15
REBUILD
UPDATE1A PROGRAM
MARC WORKING FILE (NEW)
ADD
VERIFY

Table 2. Possible Input Modes

Subsystem	Input of New Records	Input of Corrections and Verifications
Batch Tape Subsystem (Figure 1)	Off-line to tape	Off-line to tape
Batch Disk Subsystem (Figure 2)	Off-line to disk (section I)	Off-line to disk (section I)
On-Line Subsystem (Figure 2)	Off-line to disk (section I)	On-line to disk (section III)

verification of MARC records and 2) access of records by key elements in addition to the LC card number.

The major change in section I (Figure 2) compared with the tape system (Figure 1) is the implementation of a format recognition program which may be used in place of the format edit and content edit programs. All three programs will reside in the system, and either format edit and content edit or format recognition will be called in, depending on the file being processed. The output, in either case, is the same.

The update programs (Figure 2) are being designed as a series of modules which handle several transactions under the conditions that control can be taken away from one transaction before completion and another transaction executed while the first is in a wait status. Such programs are referred to as reentrant programs.*

The design work of the expanded MARC system completed to date is the result of a "first cut" analysis. More detailed study is required before the specifications become final; therefore, the programs and the organization and accessing methods for the disk files are subject to change.

---

*When processing a variety of input messages, one program may be "WAITing"—for a file action, for example—and at this time another transaction wishes to use the program. This can cause problems if the program is written in such a way that it modifies itself while being executed, or stores logic information for later use in a location other than the unique message-reference block. Programs to be used by multiple transactions in this way must be carefully written so that no logic error can be caused by this. In particular, they must not modify themselves in such a way that, when control is taken away from them, another transaction can interfere with the modification. Programs which can be entered by multiple transactions without interferences are referred to as *reentrant* programs. If a program is not reentrant, it may be necessary to have more than one copy of it in core at certain times in a multi-thread environment.[4]

PROGRAM DESCRIPTIONS

The program descriptions that follow refer to the numbers assigned to the rectangles in Figure 2.

1, 2, 3, 5. *Preedit, Fmtedit, Conedit,* and *Sort.* Perform the identical functions as described in 1, 2, 3, and 4, pages 4-5, substituting disk for tape.

4. *Format Recognition.* The process which has become known as format recognition has been described in several publications.[5] Briefly, the purpose is to remove the burden of editing (assigning content designators and fixed field codes) from the human to the machine. All the algorithms have been defined to process completed unedited English-language bibliographic data and to automatically produce an edited MARC processing formated record. Flowcharting at the coding level has been completed for the implementation of required software. During the coding phase of the project, a judgment will be made for the inclusion or exclusion of each algorithm based on frequency of occurrence versus machine processing time required to assign the content designators and fixed field codes.

The result should be an optimum balance between man and machine editing. Concurrent with the definition of the algorithms, several statistical studies were made to determine the feasibility of format recognition. The result of the analysis indicates that approximately 70 percent of the records produced will be error free, and 30 percent of the records will contain one or more errors. Therefore, not only does format recognition appear to be feasible, but it should be significant in terms of the reduction of the manpower requirements to process MARC records.

The source data for format recognition is the LC manuscript card (Figure 3)* which is the record used during the cataloging process and from which the LC printed card is produced at the Library Branch of the Government Printing Office. The data from the manuscript card will be transcribed to machine-readable form by an input typist either without any prior editing or with only a minimum amount of editing.

The manuscript card is formated into paragraphs by the catalogers, and the typist follows the following organization: she uses multiple carriage returns to indicate the end of a paragraph and a single carriage return to indicate the end of a line within a paragraph. The resulting machine-readable record processed through Preedit becomes the input into the format recognition program.

---

*The characteristics of the manuscript card will be compared with the LC printed card and the procedures will be expanded, if necessary, to include the printed card, which serves as the source data for the RECON Pilot Project.

MARC 4

SD11
.A45614 Childs, Thomas White, 1908-
no.69 Elytroderma disease of ponderosa pine in the Pacific Northwest by T. W. Childs. Portland, Or., Pacific Northwest Forest and Range Experiment Station, U.S. Dept. of Agriculture, 1968.

ii, 45 p. illus.

27 cm. (U.S.D.A. Forest Service research paper PNW-69) 0.50

NC gs 11se69 sw DO NOT SET unb 76-603289

Cover title.
Bibliography: p. 38.

1. Ponderosa pine—Diseases and pests:
2. Elytroderma deformans.

I. Title.
(Series: U.S. Pacific Northwest Forest and Range Experiment Station, Portland, Or. Forest Service research paper PNW-69)

	DDC	CRD
	634.9/08 s	76-603289
Library of Congress	69 3	MARC

Figure 3. LC Manuscript Card

The format recognition process consists of five steps:

a. Establishes the framework of the MARC processing format and initializes fixed length areas.

b. Performs the preliminary identification of the input fields, breaking the data into clusters such as the title paragraph, the collation paragraph, etc. The collation is located first, and by working back through the data an attempt is made to locate and identify the call number, main entry, and title paragraph. Analysis then continues by examining the paragraphs following the collation paragraph, identifying fields when possible. The title paragraph is divided into title, edition, and imprint; the collation block into collation, series, and price. Tags are assigned to each field

identified and the tags are recorded in the preliminary record directory. The result of this step is that the variable fields in a MARC record have been located and partially identified. For many of the fields, the three-digit tag has been assigned, for others only two digits of the tag have been established. (For example, a note is recognized as a note but no attempt is made in this step to perform the required analysis for the type of note.)

c. Analyzes each variable field and extracts information from the data to assign field indicators and subfield codes and delimiters. The analysis in this step completes the assignment of the third digit to each tag. Advantage is taken of previously processed fields through a "look back approach." A "look ahead approach" of subsequent fields is used if the field was identified in step b and the tag was recorded in the directory.

d. Performs the analysis that can only be accomplished after all variable fields are processed, i.e., where identification involves relationships between fields. For example, the first indicator in the title field indicating whether a title added entry is needed, cannot be set until after the added entry fields have been processed.

e. Completes the record assembly from the various parts of the record built during the previous processing and outputs the record for further processing. The results of the format recognition program at this point are identical with the results of the content edit program.

6, 8. *Modified Update1, Modified Update2.* Performs the identical functions as described in 5 and 7, page 5, substituting disk for tape for data base storage. The only significant difference in Modified Update2 is that records requiring modification on the master file will be copied onto the working file for action. The working file record will be modified and remain on the working file until it is verified and then it is transferred back to the master file.

7, 9. *Print Index.* Performs the identical functions as described in 6 and 8, page 5.

10, 11. *Sort, Print Program.* Performs the identical functions as described in 9 and 10, pages 5 and 8.

12. *Purge Working File.* Builds working file (new) and new accessing tables.

13. *Purge Master File.* Builds master file (new) and new accessing tables. The program for purging the master file is the same for section III and is not repeated in the schematic representation.

14. *On-line Update.* Requests for on-line retrieval and/or the modification of MARC records residing on either the working file or the master file will be made via a cathode ray tube (CRT) device. The records will be displayed either on the CRT or a receive-only

teletypewriter. The teletypewriter will be used for the display of records that require more detailed examination.

All modifications will be made to records on the working file in the on-line mode, i.e., if a record is on the master file and a request is made for that record, a copy will be transferred to the working file. Three general types of commands will be implemented:

1. A display command will result in the record being displayed on one of the two output devices. No change will be made on the record.

2. A retrieve command will request the record for modification or verification.

   a) If the record is currently residing on the working file, it will be located, read from disk to core, and a message sent to the user.

   b) If the record is not on the working file, an attempt will be made to locate the record on the master file. If it is located, a copy will be transferred from the master file to the working file and the user notified. If the record is not located on the master file, the user will be advised of this condition. (The on-line system will have access to three disk packs, the working file, the master file index, and one disk pack of a multi-disk pack master file. Therefore, it is possible to request a record from the master file which is on a pack not mounted. The request will be placed in a queue. This queue is processed in an off-line mode and the requested records copied onto the working file for modification at some later time.)

3. A modify command will modify the retrieved working file record in one of four ways:

   a) Correct: a record may be corrected by adding, replacing, or deleting data to fixed and variable fields. The modified record will be displayed on the CRT. A dialog will allow the user to further modify the record until he is satisfied, and can declare the record error free (verified) at this time.

   b) Verify: a record may be declared error free. This will cause a change in the status of the record on the working file, signaling that this record may now be transferred to the master file.

   c) Delete: all copies of this record should be deleted from the data base.

   d) Scratch: this function is included as opposed to delete, to allow a record to be removed from the working file and to have no effect on the master file. This would allow MARC editors to train on-line.

15. *Update1.5* Builds the working file (new) and new accessing tables.

ORGANIZATION OF THE WORKING FILE

The working file will contain all those records in process, i.e., not yet verified or records recalled from the master file for updating. It is anticipated that the file will not be a voluminous file. (In the current distribution service, the working file usually contains about 3,000 records.) The working file will be maintained in LC card number sequence. When a record from the working file is modified, the new version of the record will be replaced over the old version if space permits. Otherwise, the old version will be scratched and a jumpout provided pointing to the location of the new version. Only whole records will be stored on a track. Continuation records (from one track to another) will only occur under the condition that the number of characters in a record exceeds the number of bytes in a track.

ACCESSING THE WORKING FILE

The accessing method used for the working file, called Milestone Tables, is similar in structure to the IBM Index Sequential Accessing Method.

There are two levels of Milestone Tables: the first level is the cylinder table which contains the LC card number of the last record in each cylinder, and the second level is the track table which contains the LC card number of the last record in each track of the cylinder. Each table entry in both the cylinder table and the track tables refers to the next sequential cylinder or track respectively. Therefore, no information other than the LC card number is required.

Since the working file will reside on a single disk pack, and there are 200 cylinders per pack, the cylinder track will never exceed more than 200 entries. An entry will consist of three alpha bytes for the alphabetic prefix of the LC card number and four pseudo-packed decimal bytes* for the numeric portion of the number. The cylinder table will contain 1,400 bytes maximum and can be maintained in memory during operation.

The track table, one per cylinder, will be maintained in the first track of each cylinder. Since there are twenty tracks per cylinder, the track table will never exceed more than twenty entries.

The LC card number for the requested record is searched against the cylinder table to determine which cylinder the record is in. The first track of that cylinder is read, and the card number searched against the track table to determine the track the record is in. The proper track is read, and the card number is compared with the card number of each record on the track until a match occurs. If a record has been modified and the new version has not been placed over the

*A pseudo-packed decimal byte is one that has no sign.

old version, the jumpout which points to the location of the new version may necessitate an additional read.

The number of accesses required to obtain the requested record is dependent on whether the record is in the first track of the cylinder and whether a jumpout is present in the record. If the record is in the first track of the cylinder, only one access is required; if the record is in other than the first track, two accesses are required. A jumpout situation will require the third access. Since there are twenty tracks per cylinder, the requested record will reside in the first track 5 percent of the time. Therefore, on the average, two accesses will be required to obtain a record on the working file.

ORGANIZATION OF THE MASTER FILE

The master file will be maintained in date-of-last-transaction order. This is necessary for the MARC Distribution Service. The Distribution Service is designed to supply weekly, quarterly, semi-annual and annual tapes, and must include corrections and deletions to records already distributed as well as new records. The date of the last transaction will be used and a range provided consisting of an initial date and a final date of the particular distribution cycle.

All records will be added to the end of the file, packed, and when required, the record continued onto the next track. (The length of the record exceeds the space available on the first track.) When a record is copied from the master file to the working file for modification, a "change in process" flag will be set in the record in the master file. This will guarantee that a record will not be distributed to subscribers if the record is being modified and has not been replaced on the master file prior to a distribution run. In addition, any access to the master file for this record could cause a signal to the user informing him of the status of the record. When a record is replaced on the master file, the modified record is added to the end of the file. An "obsolete" flag is set in the old record plus a jumpout inserted pointing to the location of the modified version of the record. A batch-processing purge master file program will remove obsolete records periodically as required. Since this program will build a new master file and index(es) onto new disk packs, the old packs can be retained for backup.

An index will point to a record in the master file, the pointer consisting of a pack/track number and the number of the record in the track. At the end of each track a location table of two bytes per record for each record on that track will be maintained in reverse sequence, giving the actual starting position of each record in the track.

Example:

Track i	R1	R2	R3		P3	P2	P1

R = record
P = starting position of record in track

ACCESSING THE MASTER FILE

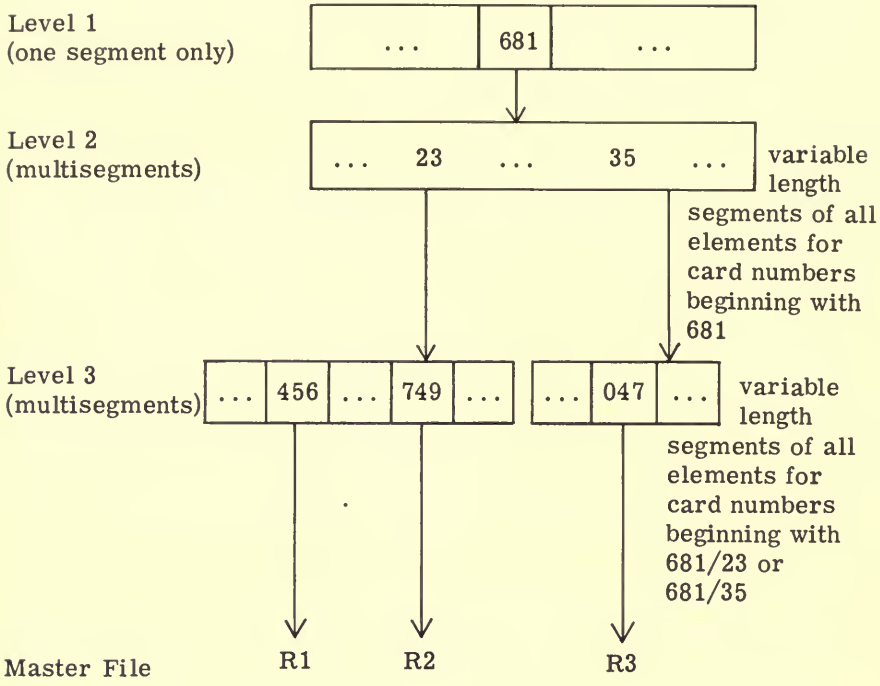
The primary access to the master file will be the LC card number. The LC card number, when expanded, is actually an eight-digit number. For example, 68-2698 expands to 68-002698. The eight digits will be divided into three parts. The first part will contain three digits, the second, two digits, and the third, three digits. The example given above would be divided as follows: 680/02/698. Since the probability is higher that all combinations of the last three digits of the card may be present, the division was made at this point. This should have the effect of shortening search time when the number is not in the index. The three parts of the number will be converted to binary to save storage space in the index; the three-digit numbers (000-999) requiring two bytes and the two-digit numbers (00-99), one byte, respectively.

The result is a three-level index and any record in the master file can be retrieved (assuming the master file pack is mounted at the time of search) in four disk accesses, including reading the track containing the record. The exception would be for those records containing alphabetic prefixes. The prefix would be considered a suffix in the index and would become the fourth level, if present. The seven-series card numbers do not include prefixes and, in the past, the number of records without prefixes far exceeded those with prefixes. Therefore, by considering the prefix the fourth level, in the majority of the cases, matching on a null prefix, i.e., a card number without a prefix, will not be necessary. The four-disk access described above and in the example below does not refer to a card number with a prefix.

The index will be a standard threaded-list tree structure in which an element in one level will point to a segment containing elements in the next level. Consider the following LC card numbers assigned to records:

Record Number	Card Number	Division of Card Number
R1	68-123456	681 23 456
R2	68-123749	681 23 749
R3	68-135047	681 35 047

The resulting index structure would be:



Each element within a segment contains information data as well as the data itself (the actual binary number portion of the divided LC card number). The information data will consist of 5 bytes assigned as follows:

Byte	Information Conveyed
1	Count of all records which can be reached from this element.*
2	Length of data portion of element (refers to byte 6 or bytes 6 and 7).
3-5	Link to segment in next lower level of index or to the master file record.

*This binary counter (1-250 and 251 meaning over 250) would equal the number of records reached if this branch of the index were traced through to completion. If this method of indexing is used when the system is expanded to allow

(Bytes 6 or 6 and 7 will contain the binary number portion of the LC card number. Whether one or two bytes will be needed depends on the level involved, i.e., level 1 and 3 require two bytes for the three-digit portion of the number and level 2, one byte for the two-digit portion of the number.)

The link portion (bytes 3-5) is constructed differently depending on whether the link is to the next lower level index or the master file record. The three bytes are considered to be twenty-four bits in length.

	Number of Bits	
Link Part	Lower Level Index	Master File Record
Flag	1	1
Pack number	1	6*
Track number in pack	12†	12†
Segment (for lower level index or record number) (for master file record in track)	10‡	5**
Total	24	24

searching by author, title, etc., this counter could serve as a guide to the user. For example, before requesting a list of books written by John Smith, the counter could be displayed and the search continued or not, depending on the resultant answer. The counter does not have meaning in the index of LC card numbers, but was included in consideration of using simpler techniques for all indexes in the system when it is expanded.

*The system is being designed projecting three million records in machine-readable form. This volume of cataloging data would require sixty-four packs. Thus, 6 bits ($63_{(10)}$) is adequate for the pack number. The one bit in the "lower level index" column is a "filler" bit.

†There are 4,000 tracks per disk pack. Thus, twelve bits ($4{,}095_{(10)}$) is adequate for the track number.

‡The minimum segment length on level one or three is six bytes plus a two byte location or nine bytes total. Therefore, the maximum number of segments per track is 7,294 bytes/9 or 810 segments per track. Thus, ten bits ($1{,}023_{(10)}$) is adequate for the segment number.

**The average MARC LC processing format record length is 633 bytes. Therefore, the number of records per track is 7,294 bytes/633 or approximately twelve records per track. Thus five bits ($31_{(10)}$) is adequate for the record number. (Allowance has been made for smaller records which will cause an increase in the number of records per track.)

Segments within a level will be variable in length. The end of a segment will be flagged with a ϕ in byte 1 (count of all records which can be reached from this element) since this byte by definition cannot contain a ϕ. If it is necessary to continue a segment onto another track, byte 6 will be set to ϕ and the bytes 3-5 (link) will point to the track where the segment continues. Segments will be packed into a track from left to right, with the same two bytes per record location table at the right-hand end of the track as described on page 21 for the master file organization.

S_1	S_2	S_3	---	L_3	L_2	L_1

S = Segment
L = Starting position of segment in track

Present thinking concerning additional indexes (author, title, etc.) includes these indexes referencing the LC card number which, in turn, will reference the master file. Thus, the system is open ended.

In the time period 1970-1971, the operation of the LC data utility will continue to be predominantly batch processing. However, we are proceeding as rapidly as possible to make on-line processing available to certain applications. The unified approach of the data utility as defined in this paper best employs the knowledge gained in developmental work and operational experience across all projects.

"Getting the show on the road" at the Library of Congress involves an effort far beyond the year 1971. Just the sheer logistics of reducing pertinent existing files to machine-readable form, as well as gearing up to include all current cataloging (approximately 200,000 titles per year) in the MARC system, will be a staggering undertaking.

If we were dealing with a simple system, we would start at the beginning and progress through each operation within LC. We would have minimal problems setting our sights and accomplishing our objectives. However, the fact is that given all the resources required, it still would not be possible to design and implement the total system in the near future. Therefore, we are attempting to place all ongoing projects and planned projects, operational or research and development, within a frame of reference, so that we have a common understanding of how we are developing and have the coordination required to do the job.

I cannot improve on Oettinger addressing this point:

> There is today a widely held point of view from which most anything, and education in particular, can be described as a collection or system of interdependent parts belonging to a

hierarchy in which a system may have subsystems of its own while acting as a mere part of a suprasystem. The process of analyzing or synthesizing such systems, called "systems analysis" for short, is touted as one of the shiniest of new technologies.

To some extent, speaking of systems is little more than appealing to a fashionable metaphor for the sake of snowing someone. Even then, this fad is not without merit as an antidote to that other pseudo-scientific fad, the precise and exhaustive analysis of an insignificant isolated effect under artificial conditions. Thinking "systems" at least reminds one that everything is related to everything else. Although always necessary in practice, ignoring any of these relations can be perilous; thinking "systems" alerts us to this peril.[6]

From the original input through the production of output (cards, MARC record, book catalogs), i.e., from the moment the book is in hand through the generation and use of the bibliographic record, each function has a high degree of dependence on every other function. In addition to this and many other technical difficulties (large character sets, complex voluminous files, etc.), we are concerned with the design and implementation of an information system over a span of time. We are planning and designing today, in terms of the potential for the future, and tied to all the work of the past. These are some of the factors that make the automation of libraries a task so difficult to accomplish.

REFERENCES

1. Leach, Theodore E. "A Compendium of the MARC System," *Library Resources & Technical Services*, 1:250-75, Summer 1968; U.S. Library of Congress. Information Systems Office. *The MARC Pilot Project: Final Report.* Prepared by Henriette D. Avram. Washington, D.C., Library of Congress, 1968, pp. 19-45; Rather, John C., and Pennington, Jerry G. "The MARC Sort Program," *Journal of Library Automation*, 2:125-38, Sept. 1969; Avram, Henriette D., *et al.* "MARC Program Research and Development: A Progress Report," *Journal of Library Automation*, 2:242-65, Dec. 1969; McCabe, Charles E. "Computer Applications in the Library of Congress Science and Technology Division." *In* American Society for Information Science. *Proceedings of the Annual Meeting.* Vol. 6. Washington, D.C., 1969, pp. 63-67; and Reimers, Paul R., and Avram, Henriette D. "Automation and the Library of Congress: 1970." (To be published in *Datamation.*)

2. Avram, *et al.*, *op. cit.*, pp. 244-49.

3. U.S. Library of Congress. Information Systems Office. *The MARC II Format; A Communications Format for Bibliographic Data.* Prepared by Henriette D. Avram and Lucia J. Rather. Washington, D.C., Library of Congress, 1968, pp. 40-44.

4. Martin, James. *Design of Real-Time Computer Systems* (Prentice-Hall Series in Automatic Computation). Englewood Cliffs, N.J., Prentice-Hall, 1967, p. 148.

5. RECON Working Task Force. *Conversion of Retrospective Catalog Records to Machine-Readable Form; A Study of the Feasibility of a National Bibliographic Service.* Washington, D.C., Library of Congress, 1969, pp. 169-82; Avram, *et al., op. cit.,* pp. 250-53; and Avram, Henriette D. "The RECON Pilot Project: A Progress Report." (To be published in *Journal of Library Automation.*)

6. Oettinger, Anthony G. *Run, Computer, Run; The Mythology of Educational Innovation* (Harvard Studies in Technology and Society). Cambridge, Mass., Harvard University Press, 1969, p. 53.

Hillis Griffin
Head, Information Systems
and Technical Processing
Library Services Department
Argonne National Laboratory
Argonne, Illinois

MARC USERS: A STUDY OF THE DISTRIBUTION OF MARC TAPES AND THE SUBSCRIBERS TO MARC

MARC DISTRIBUTION SERVICE

My paper covers two aspects of MARC: 1) the MARC Distribution Service and 2) the MARC users themselves. The MARC Distribution Service is the arrangement by which the MARC data are sent off every week to each of the users. Each weekly issue of MARC is complete on one 300-foot reel of magnetic computer tape. Thus each user receives one reel of tape containing the MARC data for that week and a printed listing showing the LC card order numbers in the shipment, their status (new, correction or deletion), and the number of new, correction and deletion records, plus a total record count.

The MARC tapes may be obtained in either of two recording modes. Most users receive the tape that is written 9-track, 800 characters per inch and is directly usable by popular third-generation computers such as the IBM 360 and the RCA Spectra 70 machines. About one-third of the users receive the tape that is recorded 7-track, 556 characters per inch. If the lower density of the 7-track tapes and the fact that more characters are required to record the same information on the 7-track than on the 9-track tapes (because escape codes must be used) are considered, then it is apparent that the 7-track tapes will always consume more length per record. One 300-foot tape will accommodate about 1,800 7-track records. The 7-track tapes are copied on an IBM 1401 system with 729-V tape drives. The 9-track tapes are copied on an IBM 360-30 system using an IBM 2401 9-track tape drives. When Argonne began the distribution program, the system had IBM 2415 tape drives and caused no end of misery. The 2415 apparently does not perform full error analysis on records written and read. This means that it can write defective records and read them back through the same device very nicely, but a 2401 or similar tape drive, with full error-checking capability, will refuse to read error records written by the 2415.

Many MARC Distribution Service users never become aware of problems occurring with their equipment until they receive some sort of input which is generated *outside* of their own computer center. Only then is it discovered that the tape drives being used are not properly adjusted to specification. This is one of the most serious problems in the effective interchange of information by magnetic tapes. One user with non-IBM tape drives runs on a small computer system and has had much difficulty reading the 7-track tapes. Testing on our 1401 and 360 equipment, we find that the tapes read perfectly well and that they are written within specification. This is of little direct comfort to the user, however, who insists that they are not readable on his equipment. His problem, although he does not want to admit it, is that *his* tape drives are out of adjustment. This tends to be verified because no other 7-track user has had any real problems reading the 7-track tapes. Since discarding the 2415 tape drives we have had no reported problems with the 9-track tapes, and the problems seem to be almost nonexistent except for the user with the low-cost tape drives.

Another procedure which we follow with the tapes is that of cleaning each reel. We feel that dirty tapes might have caused some early problems, and in fact we normally clean new tapes before introducing them into our system. One might assume that brand new tape, fresh from the supplier, would be without peer for accuracy and flawlessness. I think that this was probably so in the case of second-generation equipment that used 7-track tapes. The third-generation equipment, however, writes nine tracks on the same physical tape simply by using more of the tape surface from edge to edge. This means that data are being written closer to both edges of the 9-track tapes than on the 7-track tapes. It follows, then, that one needs more complete contact of the tape with the read-write heads from edge to edge.

Consider a new piece of magnetic tape. It has been slit from a larger sheet of coated mylar and spooled, usually without further cleaning. There is some debris remaining on both edges of the tape from the slitting operation. This debris will tend to lift the edges of the tape as it travels over the read-write head, and will prevent proper recording at constant amplitude across the full recording surface. We have removed a lot of junk in this cleaning operation. Sometimes, in fact, we will even clean a new reel twice because of the high level of debris accumulation from the first pass. A tape cleaner has become a necessity in the computer center, if for no other reason than to enable the cleaning of the tape upon which the master file is written since the computer may reject the tape because of errors probably caused by dirt.

Argonne obtains the tapes on a bid basis. We ask tape manufacturers to bid on specification and quantity. Our first 1,000 reels cost

$4.75 per reel, but we had bids up to $7.00. The next round brought a low bid of $4.25 per reel, including a plastic can and a more sturdy shipping container, and this price seems to be holding. All of the tape we use is new and is certified full width at 800 characters per inch (1,600 fci). One of the recurring questions raised by the users concerns a possible swapping arrangement for used tapes. Users would like to have the MARC Distribution Service take back their old tapes for credit and reuse these tapes in the distribution cycle, such as was done in the MARC I program.

Reuse of tape introduces bookkeeping problems and additional expense in the operation. Since the environment in which the tape has been used and stored is unknown (and in some cases, unbelievable), the distributor would have to clean and recertify each reel of tape simply to assure himself that the new records written on the tape for shipment to another user would be readable. These costs would approximate the cost of new tape and would consume machine time which is usually not available. I realize that the accumulating supply of used MARC tapes in libraries around the world has potential for environmental pollution which I do not want to consider here. Perhaps when the first MARC tape is dredged from the mud of Lake Michigan or Lake Erie, we will be forced to consider other ways for their disposal. I might say that we regularly have requests from the programmers in our accounting department to sell them these mini-reels in dozen lots. They find that the reels are ideal for program testing and the storage of small files of data. In any event, I feel that the reuse of MARC tapes in the MARC Distribution Service is not justified at present cost levels.

The MARC Distribution Service consists of three programs written in PL-1 and assembly language. One program writes a user list, tape-reel labels, and address labels for the tape shipments. The user file is maintained on punched cards because users change their addresses frequently. The reel labels and address labels are printed in sets—7-track and 9-track—for convenience, but the user file is maintained in strict alpha sequence, also for convenience.

The 9-track tape that we receive from the Library of Congress is in an interim format. It is an IBM 360 form "V" record with the control number (the LC card number) in EBCDIC in a leader record of fixed length, to which is appended the MARC ASCII record. The use of this interim format lets us handle the MARC record rapidly in the distribution programs without having to translate character-form ASCII record lengths to EBCDIC binary form record lengths to tell the channel the record length for data transmission to the output device. It also lets us block these records. The second program in the distribution reads these intermediate MARC records, makes blocked form "V" records of them, picks up the LC card number and puts it into an index page array in core, and then passes the form

"V" records (now blocked) to a 2311 disc. The index pages are dumped to a spool disc as the page array is filled, page by page. At the conclusion of this phase of the program a second phase dumps the index records from the index spool to the printer at write speed in the number of copies called for in the program load card. This provides sufficient index copies (plus some overrun) for all 7- and 9-track users, and for stock.

The next program writes form "U" records on tape for distribution. Since we only have two tape drives, it is most efficient to carry the master file on disc, blocked (to conserve channel time and reduce the number of input/output calls), and to unpack the form "U" MARC record from the intermediate, write it on the two tape drives, and then call for the next input record, which is probably already in core and needs only to be deblocked. This makes the operation move very rapidly. It moves so rapidly, in fact, that the program is generally either rewinding the completed reels in low-speed rewind (it never copies enough tape to get into a high-speed rewind), or waiting for the operator to demount the completed tapes and mount new blanks for the next copy cycle. This is where the time really goes—into demounting and mounting the tapes. The copying goes very rapidly, so rapidly in fact that sometimes the operator has not had time to get the last tapes into the cans and the next tapes out of the cans.

After the copying has been completed the indexes are decollated and returned with the tapes to the library. We retrieve the 7-track tapes from the other computer center where they have been copied on the IBM 1401. We box the reels and indexes, label the boxes, and ship them out. The shipment generally gets to the post office air mail facility at O'Hare Airport during the afternoon or early evening. Some users receive delivery of their tapes the next day. O'Hare is an excellent shipping point since there is a flight to almost anywhere in the world leaving within any given four-hour period during the day or evening. Even when problems have arisen and the tapes have not reached O'Hare until Sunday afternoon, no one on this continent has failed to receive Monday delivery.

I might emphasize that this is not a profit-making operation for us. When the MARC Distribution Service was started, consideration was naturally given to who was to distribute the tapes. The Library of Congress felt that their computer capacity was already sorely overtaxed, and that it should be done outside the library. One method was to go out on bid for a contractor. The problem at that time was that no one knew how many subscribers would eventually materialize, what the ratio of 7-track to 9-track subscribers would be, and so forth. The distribution programs also had to be written. Since we had been involved in MARC for some time, we had some interest in seeing that MARC II got off to a good start. Therefore Argonne acquiesced when the Library of Congress asked us to take on the

distribution job, including the preparation of the necessary programs. We assumed the job with the understanding that, when the system had stabilized, we would expect the library to find an appropriate contractor to carry on or else take it under their own roof. I should comment that before we had sent out the first tape, the pessimists had real doubts that we would have more than five subscribers. The optimists felt that we would *surely* have at least ten subscribers—maybe even more. At the end of the first year of MARC II we had nearly eighty subscribers in the United States and abroad. MARC II was truly an international system, and probably the largest system for information distribution in existence.

As we enter the second year of MARC II we have seen a rather dramatic decrease in the number of subscribers to the system. I do not attribute this to MARC as much as to a failure on the part of the library community to become actively involved in using the computer in their own library operations. The number of subscriptions has continued to rise, although not to the number of first year subscribers.

THE MARC USERS

There exists a MARC users group within the hallowed confines of the Information Science and Automation Division of the American Library Association which meets in mid-winter and in mid-summer for discussions, and is one of the most listener-oriented groups that I have ever encountered. Very few people who attend the meetings are willing to say what they are doing with the MARC tapes they receive. The users group has served as a forum and spawned a committee that is attempting to get some research going on a search key. I believe this committee does have potential for bigger and better things, and I think that this must come to pass.

In November 1969, Lawrence Leonard of the Library Research Center at the University of Illinois sent a questionnaire to the MARC users through the American Library Association. A follow-up went out in December. The survey went to seventy-four users, forty-six (61 percent) of whom replied. Of those who replied thirteen (28 percent) indicated that they were using the tapes. It is probably safe to infer that most actual users were among the respondents, and that few users did not respond. This would lead to the estimate that only 18 percent of MARC II subscribers were making any real use of the tapes, or were close to making any real use. The survey indicated that, of those who replied, the following were actually using or planning to use MARC in the following ways:

	now doing	planned
Accessions list	2	3
Acquisitions	4	12
Bibliographic searching	2	0
Book catalogs	3	20
Bibliographies	1	4
Book processing (cards, labels)	1	6
Book selection	5	5
Catalog cards	3	21
Cataloging	1	7
SDI (selective dissemination of information) system	2	6
Data base		2
Information retrieval		1
Training (library school)		1
Do not know		1

The return is not really sufficient to be statistically significant, but it is interesting. A short time ago I took my own informal poll of some of the MARC users known to me personally. This was a telephone poll, so there was an opportunity to follow whatever line of inquiry developed. Some of the comments that I received from non-users included these:

No access to a computer but want the back files
Working actively on it
Other problems in-house (circulation, etc.)
Not enough priority to get computer time
Don't know what to do with it (i.e., MARC)
No programmer in the library
Can't get programming done
Communications:
    Librarians don't know anything about the computer
    Computer people don't understand "our" problems
No support

At some institutions the weekly tapes are apparently checked in like journals, shelved and never used. We have sent letters to some subscribers, advising them to return bad tapes, or notifying them of mistakes, etc., and asking for a reply. We have never heard from some of them. The users who receive the tapes and do not use them ignore the problems of print-through on the tapes which they file on the shelves, expecting to pull them out sometime and use them. They also ignore the economics of having to merge (or better, trying to merge) all of those weekly issues when they could obtain a cumulated tape, for the same amount of money, when they are ready to begin in earnest with their MARC applications. One can only wonder why a

library would subscribe to the MARC tapes if there were little or no prospect of using them. I can only conclude that there is probably some prestige associated with being a subscriber library.

If we look at the computing power available to the respondents, it looks something like this:

IBM 360 (mostly model 40 or 50)	33
IBM 7094	2
Univac 1108	2
Burroughs B5500	1
CDC 6600	1
PDP/10	1
RCA Spectra 70	1
Sigma 7	1

This rather interesting exhibit indicates that the large majority of respondents have IBM 360 hardware available to them, and that most of *this* group have rather substantial units available to them (model 40 or model 50). The only real difference between these two subgroups is that the model 40 users are probably using the disc operating system. None of the respondents are using the IBM 1401 as the basis of their system, contrary to expectations.

What can be done about this problem? How can libraries use the MARC tapes to help themselves? One solution in which someone else does the work is the sharing of computer programs among libraries. It would appear that the group of 360 users would be in a fairly advantageous position in this regard. The problem arises in trying to fit programs written for the larger machines onto the smaller machines. There generally tends to be rather good upward compatibility, but rather poor downward compatibility. What may be needed is some device for announcement and description of programs which users could make available to others.

This raises the problem that most librarians do not know a program from first base, and are totally incapable of assessing the relevance of the announced program to their operation and computer installation. I can sympathize with those who, having gone to the trouble and expense of generating their own program, would refrain from general announcement simply because they do not have the time to deal with the flood of unenlightened inquiries that will descend upon them. I have received some unbelievable responses from librarians when I asked them what kind of computer was available to them. (One had a computer called a Fortran, for example.) I do not think that it is the program writer's responsibility to provide an education in library data processing to each requestor. Librarians should take it upon themselves to learn the vocabulary, the concepts, and become proficient in this area if they expect to deal with it. I am very frankly

shocked at the lack of competence of many librarians who profess to have responsibility for the application of computer technology to their libraries. They seem to view themselves as middlemen between the library and the computer people, but they seem to have little comprehension of the world of the computer. They are too proud to learn programming and too dignified to learn about the hardware. This is the sort of knowledge that is not taught at one-day or three-day institutes and there is no easy method of "instant enlightenment." Without this knowledge their title of "automation librarian" is an empty one and the advice they provide their institution with may be dangerous in its ignorance of the real problems of the situation. Our library schools make very little contribution to the solution of this problem.

Another solution is a series of ready-made systems which could be purchased by any library. The systems might be available in the form of system flow charts or as program packages. The problem here is that the system should be tailored to an individual library and to the computer facilities available to that library. It would be folly to write program packages for large computer systems because this would preclude their installation on the many available smaller systems. The smaller systems, then, present individual constraints system by system in the availability of direct-access files, tape drives, printers, print trains, and a host of other problems. By the time a system has been designed for the lowest common denominator it really is not worth much to anyone. There is a great deal to be said for systems that are responsive to the needs of those who use and support them. Few off-the-shelf systems meet this need.

Another solution absolves the librarian of any responsibility for the system and guarantees a lack of contact with the computer. This solution is the use of a service bureau and its programs for processing the MARC records; it is somewhat analogous to centralized processing. The service bureau, using its set of programs, processes the MARC records in specified ways, giving the libraries specified output. The NELINET operation, the planned OCLC operation, and the printed card service of Richard Abel and Company are examples of such applications. This enables libraries without computers to obtain the benefits of the MARC record through cost-sharing with other libraries. While they may not obtain a customized and tailor-made product, they do obtain a usable and helpful product which they might not otherwise have. We are planning to share our MARC selection and catalog information programs with a group of college libraries, for example. We hope that, if the program is successful, we may see some kind of centrally based catalog information system, operated jointly by these users for themselves without our intervention. We feel that this sort of centralized processing can be very attractive to smaller libraries, but that libraries with access to

adequate computer facilities should be encouraged to do things for themselves, simply because they can do *more* things for themselves on their own equipment.

This is the cloudy picture as it appears today. Those who are using MARC are using it because they have rolled up their sleeves and done something to solve their problems. I think they should be given a lot of credit. The MARC record is an extremely sophisticated animal, and is extremely difficult to manipulate properly if one has never done this sort of thing before. It is new ground for most programmers, but we are now beginning to see more sophisticated records in many of our data processing applications as we move toward more complex information systems. The MARC format is, in a way, ahead of its time, but it is eminently usable and manipulable for any products which can be defined. As we find more people who are willing to get in and work with it, we will find more successful users. There is really no other road to success than through hard work.

R. E. Coward
Head of Research and Development
The British National Bibliography
London

THE BNB MARC PROJECT

> The MARC project has progressed from a pilot to test the feasibility of a distribution service of centrally produced machine-readable cataloging data to a full-scale operational system in the design stages in two short years. The library community, both here and abroad, has accepted MARC and recognizes its potential for the future. The single most significant result of MARC has been the impetus to set standards.[1]

These sentences were written by Henriette Avram as the opening paragraph of the MARC Pilot Project final report. By the time that report appeared in print, the potential of MARC had been fully realized in Great Britain, and a remarkable chapter in the history of Anglo-American cooperation had been written. This paper is an analysis of the British achievements over the last two years, an appraisal of the philosophy behind the British MARC Project and an examination of the suitability of MARC as a basis for future progress.

Before tackling these themes I would like to acknowledge the enormous contribution that was made to the British MARC Project by an American librarian whose name is familiar on both sides of the Atlantic but not perhaps associated with automation. The solid foundation of cooperation between the Library of Congress and the British National Bibliography (BNB) was laid on the American side by John Cronin who opened up the dialog between the two organizations and taught us all that the art and science of cataloging is the same on both sides of the Atlantic. John Cronin built the bridge we used for MARC.

MARC NATIONAL AND INTERNATIONAL

The British and American MARC Projects are shaped by many factors, most of which are rather more significant than the amount of money made available by our respective governments to carry them through. The enthusiasms, the ideas and perhaps the style of a very small group of people are decisive influences; however, much care is taken to secure the agreement of the profession as a whole to the actual course of development. The immediate environment of the project can determine the short-term course of events, particularly

when the ultimate goals are uncertain and when intermediate objectives are determined by economic factors. To put the matter bluntly, the course of the United States MARC Project cannot veer significantly from the priorities of the Library of Congress, and the British MARC Project must fit broadly into the future pattern of development of the British National Bibliography. Those of us who feel that pure logic rather than practical judgment and practical compromises must determine the course of these important projects should direct their attention not to MARC, but to the national bibliographic policies of their countries—if indeed these policies can be identified.

MARC, for all its much vaunted flexibility (which is purely superficial), must fit into an existing pattern. In a general sense this means that the separate halves of our joint project must develop along lines that are first of all acceptable to a local institution environment, which is in its turn acceptable to a national environment. The level of communication between national systems must be based on an understanding of these factors. This includes an understanding of the relative levels of compatibility and the difference between local, national and international standards. Without this understanding our efforts to create an international MARC network will be frustrated by a lack of communication between ourselves as librarians long before we can tackle the much easier problem of communication in machine-readable form. By concentrating in this paper on the particular factors that have so far shaped the development of the British MARC Project and will shape its future policy, I hope to identify more precisely the wide areas of agreement with the Library of Congress Project and the areas in which agreement must be urgently sought.

BRITISH MARC

The British MARC Project has been created in an environment which is significantly different from that enjoyed, or endured, by the Information Systems Office at the Library of Congress. It is a pity that nobody has had practical experience in both systems—a comparative view from the inside would be a fascinating document. However, I have benefited beyond measure over the past two or three years from the Library of Congress's "open door" policy.

On the basis of this limited experience I have formed the view that the British MARC Project enjoys several advantages which are denied to the Library of Congress Project. There are also serious disadvantages, but on the whole the Library of Congress environment appears to force a particular line of development that will become increasingly difficult to fit as a total package into an international context. The British situation is very different indeed. In Great Britain there is now a total physical separation between the national bibliography (BNB) and the national library (British Museum). The

British National Bibliography concentrates its entire attention on the provision of published bibliographic services and as such it is a machine for producing bibliographies. From the moment it was decided to proceed with the MARC Project, it became obvious that this bibliographic machine would be taken to pieces and re-built around a computer. We still talk about MARC as an experimental project; however for BNB the experiment ended a year ago. MARC is a production system that must work and BNB must now become both the producer and the biggest single user of the MARC record in Great Britain (a situation which concentrates the mind wonderfully). Most of the professional librarians on the staff of BNB are directly involved in the rapid changes that are taking place. In the past year the MARC Project as a separate organization has been getting smaller and smaller; there are now only two keyboard operators, two programmers, two librarians and two planners. Everyone else is handling MARC with one hand and the mutilated processes of BNB with the other. Next year a small team will concentrate on the external features of the MARC Project and the rest of the MARC organization will simply be the new British National Bibliography.

The status of BNB as a national bibliography is rather unusual. The council of the British National Bibliography is a private, non-profit organization established to provide bibliographical services; BNB is totally independent and totally self-supporting. Its only income apart from the MARC Project grant is derived from the sale of the services it provides. This is a rather extreme situation to be in and perhaps not one to be recommended for the future. However, it does bring BNB right up against one of the long-term objectives of the MARC Project, i.e., its practical use in the production of more efficient, more economic bibliographical services. The first objective for BNB must be to utilize MARC and, moreover, to utilize it in a highly cost effective manner.

It has often been remarked that the introduction of a computer system provides a unique opportunity to carry out a complete re-appraisal of the services provided by an organization. In the context of MARC it is nowhere more true than in a national bibliography. There is virtually no hangover problem from the past. National bibliographies cannot afford to change their spots too often, but after twenty years (which is the life span of BNB) change is necessary to keep up with changing conditions—MARC has provided this opportunity. It has also provided the means to start in a small way a process of rationalizing, integrating and improving bibliographic services in general. A separate and independent bibliographic organization cannot afford to deploy more than a very small part of its resources on work that is not immediately productive. Long-term commitment to developing standards or new systems is almost out of the question. The MARC Project has provided the impetus, if not the

resources, to take a wider view. During the two years that BNB has been developing the British MARC Project, much more fundamental activities have been going forward. These include bibliographic integration with book trade services, changes in the design of the national bibliography and the development of a new generalized indexing system.

MARC AND THE BOOK TRADE

In Great Britain the book trade was getting ready for computerized book ordering and cataloging systems as early as 1966 when the Publishers' Association commissioned a report on the need and feasibility of a standard system of book numbering. It was quickly appreciated that a purely book trade numbering system would not satisfy the large library field, and the initial proposals were expanded to a plan for total numbering. Less than three years after the scheme was approved there is a system which gets numbers printed on over 90 percent of all new books, and numbers can be allocated through a central agency for all the remaining books. This operation is run by Whitaker & Sons, who produce BNB's book trade bibliographies. BNB's function is to notify the standard book numbering (SBN) agency of any item received without a number. In 1971 this will be provided before the book is entered on a MARC file, and before it is entered in BNB or the trade bibliography. In Great Britain total book numbering is an accomplished fact. The next stage, already well under way, is an international numbering system. The implications of this development on the United States will be considerable. There is already a standard book numbering agency in New York with an impeccable list of sponsors but is there any commitment by the agency and the Library of Congress to establish 100 percent coverage? There is no need to worry about the great mass of trade books. In their own interests the trade publishers, the trade associations and the trade bibliographies will quickly establish an efficient system. The non-trade material is a library problem and a library responsibility. It must be brought into the system as it enters the Library of Congress. Will it be, or does this conflict with other priorities? This situation is a classic example of a conflict between a good national system—the Library of Congress card number—and a better international system—the International Standard Book Number. Both can run side by side but the international system must be vigorously implemented.

With the SBN system established, the book trade bibliography in Great Britain was quick to see the potential of MARC. The cooperation between Whitaker & Sons and the British National Bibliography was extended to the field of data processing. Last year BNB and Whitaker developed what now appears to be a highly efficient MARC system for producing the trade bibliography (see Figure 1). This

project from design to full operation took exactly eight months. As a result the printed book trade services have now been dramatically improved and by April 1971 there will be in existence and use a massive machine file of about 300,000 items representing British books in print. This is a genuine MARC compatible system. In the spring of 1970 the final touches were put to the program which translates the book trade record to a standard communications format using the BNB MARC implementation. Already several large book-purchasing organizations in England are planning to use this file as a book order data base with records available as soon as the book is advertized. This project fills an ugly gap in MARC-based systems by providing at the time of order, however early that may be, a record which carries adequate temporary cataloging information until it is replaced by a full MARC record from BNB.

MARC AND THE NATIONAL BIBLIOGRAPHY

It may seem an unusual set of priorities that puts the conversion of another bibliography to machine-readable form above our own. The fact is we have never been seriously concerned about the data processing aspects of the change-over. I would almost go so far as to say that creating catalogs from a MARC record and using computer typesetting techniques can be considered to be a perfectly routine operation. In England the most difficult problem is finding a suitable Photon or Digiset or whatever is to be used.

It is, however, a very different matter to decide exactly how a bibliography should be organized. There is, after all, an almost embarrassing number of choices in a MARC record.

The exercise of trying to design and produce a first-class national bibliography from MARC is one of the finest ways of testing the theoretical soundness of the system. In many ways MARC stands up fairly well to the test. It has been criticized by some who did not fully appreciate what MARC is trying to achieve as being too complex: this is nonsense. The analysis that has been carried through in the British and LC MARC implementations falls short of what is really needed. Other problems include an inadequate title breakdown, series fields which are messy and illogical, notes which are inadequately specified, and added entry situations which have not yet been fully analyzed. Sometimes we find ourselves checking a program specification to decide exactly how to input a particularly complex record, a sure sign that the initial design is inadequate. In making these criticisms I am of course referring to the British MARC implementation, not the Library of Congress implementations. However, except in one or two instances, the British record identifies specific subfields that are grouped as a broader subfield in the LC record.

We have been criticized, and rightly so, for making unnecessary

PUBLICATIONS OF THE WEEK

This list is compiled by J. Whitaker & Sons Ltd., and is copyright. It consists of books published or re-issued, given alphabetically under author, title and subject where this forms part of the title. Dates of publication are given as (1.70) for January, 1970, and prices are net unless marked *. Main entries are shown beginning in bold type. Standard Book Numbers are added to each entry by courtesy of The Standard Book Numbering Agency, Ltd.

The list is also published in CLASSIFIED form monthly in *Current Literature*, and at quarterly intervals in both classified and alphabetical arrangements: THREE MONTHS (Jan. to Mar.) in May; SIX MONTHS (Jan. to June) in August; NINE MONTHS (Jan. to Sept.) in November and the ANNUAL VOLUME (cloth bound) in April, 70s. net net. The four issues of *Whitaker's Cumulative Book List* are supplied post free at 5gns. per annum. *The Bookseller* monthly alphabetical list is now published separately as *Whitaker's Books of the Month and Books to Come*, covering a period of three months, 12 issues a year, £6 per annum post free.

The symbol "n.i." indicates a new issue, cheap edition or reprint at a new price

Please send notices of omissions or corrections as early as possible to the Publications Editor to ensure inclusion or correction in subsequent cumulative issues.

A. B. M. *See* Antiballistic Missile.

Abolition of the Brazilian Slave Trade: Britain,Brazil and the Slave Trade Question,1807-69 (Bethell) D8.xvi,426. 2M. Latin Amer.Stud. 90/ Camb.U.P. (3.70) Pol 521 07583 1

About Garlic: The Supreme Herbal Remedy (Binding) C8.64. sd. 5/ Thorsons (3.70) Med 7225 0148 X

About Soya Beans (Binding) C8.64. sd. 5/ Thorsons (3.70) Med 7225 0147 1

Absalom and Achitophel (Dryden)
Ffol.77. facsim.e. 35/ Scolar P. (2.70) Poe 85417 219 X
Ffol.77. facsim.e. sd. 12/ Scolar P. (2.70) Poe 85417 218 1

Accounting, Introduction to the Theory and Context of (Sidebotham)
C8.224. 2r.e. C.I.L. S. 40/ Pergamon P. (3.70) Com 08 015620 7
C8.224. 2r.e. C.I.L. S. sd. 30/ Pergamon P. (3.70) Com 08 015619 3

Ackoff, R.L. Concept of Corporate Planning. D8.172. Ill. 70/ Wiley (3.70) Com 471 00290 9

Adams, Charles F & Adams, Henry. Chapters of Erie and Other Essays. D8.iv,428. n.i.of 1827e. 93/ Kelley,U.S.A.: David & C. (2.70) Hist 678 00220 7

Adams, Henry. *See* Adams,Charles F & Adams,Henry.

Adolescent Behaviour, Studies in: Project YG,Youth in Goteborg (Andersson) Tr.fr.Swedish A.Read. R8.400. tabs. sd. 70/6 Almqvist (1.70) Soc 85132 004 X

Advanced Guide to Coin Collecting (Linecar) D8.240. 60/ Pelham Bks. (3.70) Art 7207 0224 0

Advances in Radiation Chemistry.
v.1. Ed.M.Burton & J.L.Magee. M8.412. Ill. £7/15/ Interscience (3.70) Che

Aesthetics, Four Stu
sC8.120. 57/
sC8.120. sd.

Africa,East. *See* Ea

African Rural-urban Migration: Movement to Ghana's Towns (Caldwell) R8.xi,257. 95/* C.Hurst (12.69) Soc 900966 34 3

After Slavery: Negro in South Carolina during Reconstruction,1861-77 (Williamson,Joel) IC8.454. n.e. Chapel Hill Bks. sd. 34/ Univ.N.Carolina P. (3.70) Hist 8078 4047 5

After Utopia: Decline of Political Faith (Shklar) D8.322. n.e. sd. 26/ Princeton U.P. (3.70) Pol 691 01966 5

Agrarians,Southern,as Social and Literary Critics: "Tillers of a Myth" (Karanikas) M8.250. n.e. sd. 23/ Univ.Wisconsin P. (3.70) Lit 299 03974 9

Analysis of the Phenomena of the Human Mind (Mill) 2v. D8.xxi,453: vi,404. n.i.of 2e.(1869). set £11/11/ Kelley,U.S.A.: David & C. (2.70) Psy 678 00330 0

Analytical Chemistry in Space (Wainerdi) C4.284. 65ill. Mon.in Anal.Chem. £7 Pergamon P. (3.70) Chc 08 006887 1

Analytical Chemistry of the Polyurethanes (David & Staley) M8.640. Ill.. High Polymers S. £13/10/ Interscience (3.70) Che 471 39293 6

Anatomy, Living (Lockhart) R8.96. 214ill. n.e. sd. 15/ Faber (3.70) Med 571 09177 6

Ancient Britain. sD8.62. Col.ill. Jun.Ref.Lib. bds. 6/6* Macdonald (3.70) Schl 356 03234 5

And Now it's Sables (Spanier) D8.192. Ill. 30/ Hale (3.70) Bio 7091 1240 8

Anderson, Adam. Historical and Chronological Deduction of the Origin of Commerce,etc. 4v. C4.lxxviii,2429. n.i.of 1801e. £61/19/ Kelley,U.S.A.: David & C. (2.70) Pol 678 00259 2

Anderson, Charles P. Lettering. D4.176. Ill. 94/ Van Nost.Reinhold (1.70) Art 442 11288 2

Andersson, Bengt-Erik. Studies in Adolescent Behaviour. Project YG,Youth in Goteborg. Tr.fr.Swedish A.Read. R8.400. tabs. sd. 70/6 Almqvist (1.70) Soc 85132 004 X

Andreski, I. Old Wives' Tales. D8.176. 35/ Routledge (3.70) Nat 7100 6655 4

Andrews, John B. *See* Commons,John R & Andrews,John B.

Antiballistic Missile (Ed. Chayes & Wiesner) D8.304. 42/ Macdonald (3.70) Eng 356 03103 9

Anticipations,Uncertainty and Dynamic Planning (Hart,Albert G.) D8.xiii,98. (2.70) Pol 678 00040 9
Hutchinson (3.70) Fic 09 101090 X
ns in the Theory of Value
.100. n.i. 57/ Kelley,U.S.A.: David & C. (2.70) Mat 678 00082 4

Figure 1. Trade Bibliography Produced from the MARC System

Arbon, M.R. *See* Carter,R.F & Arbon,M.R.

Art. *See Also* under Country.

Art, Theories of Modern (Chipp) D8.300. n.e. sd. 47/ California U.P.: I.B.E.G., 2-4 Brook St.,W1Y 1AA (2.70) Art 520 01437 5

Arterial Disease, Peripheral: A Physician's Approach (Richards,R.L.) D8.126. 40ill.7tabs. 50/ E.& S.Livingstone (2.70) Med 443 00692 X

Arts and Crafts, Junior: Annotated List. D8.20. sd. 4/ (members 3/) Nat.Bk.League (2.70) Bib 85353 039 4

Ash, P. Nurse for the Season. C8.192. 14/ Mills & B. (3.70) Fic 263 51413 7

distinctions here and there. This redundant information can be easily weeded out—if everyone would agree that it is redundant. I am much more concerned about the distinctions that are not made. To illustrate this point let me cite one fairly common problem which we will have to handle by extending MARC coding. If we are creating author index entries, it is extremely useful to know whether the author statement following the title needs to be included. If it is a joint author we want to leave it in, if it is a single author adding nothing to what has already been stated, we want to drop it. These considerations are important when massive indexes are being printed and we will probably need to create a distinction. This type of problem is perhaps fairly small and insignificant, and I am inclined to take the pragmatic view that the public MARC—that is, the exchange tape implementation—should be allowed to settle down now for a year or so while everyone gets some experience. MARC should not be like the Library of Congress classification with new extensions just arriving every now and then and which can be safely ignored. It is rather more difficult to ignore changes in a MARC record. In the meantime BNB will continue to develop an internal implementation that will certainly have many features that we will quietly remove before the exchange tapes are produced. A clear distinction will be made between the local BNB implementation and the exchange tape implementation. This is a logical development. We all accept that the machine structure of the exchange format is a carrier wave between machine systems. Should not the record also be considered as a carrier wave that is logically distinct from the implementation standards of the producer and the receiver?

For every MARC record there are three distinct categories of users: 1) the actual producer of the record—at present the Library of Congress and the British National Bibliography, 2) the local user within the country of origin, and 3) the international user. In attempting to fix implementation standards it will be impossible to face three ways at once. The conflict between the producer and the local user within the country of origin is no problem. The conflict between the user in the country of origin and the international network will become serious. Taking my previous title field example, if all British MARC users demand a more complete analysis, then we will have to provide it regardless of any agreements made with the Library of Congress. The Library of Congress is in exactly the same position.

The solution to this problem is not difficult. In fact much of the work leading to a solution was done by Ann Curran and Henriette Avram three years ago when they produced "The Identification of Data Elements in Bibliographic Records."[2] At the international level we should stop looking at the BNB implementation or the LC implementation. All we need is a conceptual interface between implementations—not a real subfield code for subfield code identity.

Translating a format through an interface is a fairly trivial data processing operation. Using this approach we can concentrate on establishing the minimum level of analysis and the minimum set of fields that we will call a MARC record, and the optimum sets that might be achieved between two closely related systems such as BNB and LC.

The above diversion was caused by the remark that BNB would need to add refinements to the record analysis to produce a national bibliography. When this is done the British MARC Project will have transformed the two major bibliographic services in Britain. The improvement in the services provided to librarians and the book trade will be dramatic.

The new national bibliography will adhere rigidly to standards where they exist. The classification standard will be Dewey—probably the eighteenth edition to begin with. The cataloging standard will be the Anglo-American code (British version). Unfortunately there is no filing standard and as yet no one has given any serious thought to a subject indexing standard. In the existing MARC record no one has given any serious thought to subject indexing. This is an omission that must be remedied and as BNB will be using MARC records within a few months to produce a classified catalog with a specific index, a great deal of effort will be concentrated in this area.

A MARC SUBJECT INDEX

Anyone wishing to construct a classified catalog from the existing MARC record has a choice of two classifications, both of which are non-specific. There is no indexing provision at all; obviously MARC was born in the home of the dictionary catalog. Faced with this situation, BNB had the choice of indexing outside MARC or devising a special indexing system which would be carried on the record. Our final choice has been a compromise, a string of indexing terms will be carried on each record. This will be backed up by a machine-held thesaurus which will hold the necessary framework of cross-references. In planning this system BNB is breaking its rule that each MARC record should be totally self-contained; however I hope it will not be long before MARC subscribers have direct access to the master store of records and ancillary files. Large storage systems with direct links between local computer and central data banks are a rapidly developing area of technology.

The indexing system BNB is developing has two unique features: 1) it is totally hospitable and can be used to index any subject at any depth of analysis and 2) it can also be used to index back into any classified arrangement. These factors are essential in the context of MARC. The index can best be regarded as a bi-product of a new classification technique now in use at BNB. The history of classification theory shows a steady progression from the Dewey concept of a

universe of knowledge subdivided into smaller and smaller subject disciplines, through Bliss and Ranganathan to the special-faceted schemes which allow a great freedom in the combination of concepts taken from various parts of the system.

BNB has now taken the final logical and evolutionary step of starting the process of classification by identifying the individual concepts in a compound subject and expressing their relationship to each other in a symbolic form. The result is a subject statement which is absolutely neutral and precise and can only have one possible number in a formal classification schedule, even though the symbolic notation in the schedule will not always carry the same depth of analysis. A particular subject statement, or concept analysis, is permanently associated with a unique set of alternative system statements. At BNB there are already four such systems—DC, LC class number, LC subject heading and a MARC subject descriptor string. The subject descriptor string is the indexing statement that will appear on British MARC records later this year. It will be used by BNB as a rotated element index, but it could be used equally well as an uncoordinated keyword system. At present the system is being tested on general intake at BNB and in one or two highly specialized areas. An example of depth indexing in the field of sociology of education is given in Figure 2.

So far I may have given the impression that in Great Britain we do not primarily think of MARC as a method of exchanging bibliographic data and this is partly true. We think of MARC as a data base which is being created to form the foundation of bibliographic services of the new decade. In England, as in the United States, bibliographic services are provided at the center and these services are utilized and added to at the local system level. Experiments in local handling of MARC are going forward but we feel that at this stage such projects should remain strictly experimental. There are several reasons for this; the immediate one being that we still need time to evaluate what we have learned and to implement changes. Maybe libraries are anxious to start utilizing MARC data but they would be well advised at this stage not to go too far. The more important reason is that any library system using MARC today is almost inevitably assuming that a MARC service is just another centralized service which it can purchase like a catalog card; nothing has really changed. There is absolutely nothing on a MARC record that has not been fully available for fifty years or so on the Library of Congress card. So what is all the fuss about? The plain fact is that it is the computer that is changing the bibliographic framework within which a library operates. MARC might be a trigger mechanism for local change but it really cannot be anything else while we limit ourselves to distributing in machine-readable form what was already produced in visual form.

In England we can perhaps enjoy the illusion of progress and

Turnover rates. Higher education. United States
Staff turnover rates. Influence of organisational size & organisational complexity 033
Typologies
Education systems 086

U.S.S.R. *see* **Russia**
Uganda
Educational development *compared with* Educational development in Kenya 046

Underdeveloped countries
Related headings:
Newly independent nations
Undermanning
Related headings:
Teacher undermanning
Underprivileged children *see* **Socially disadvantaged children**
Union of Soviet Socialist Republics *see* **Russia**
United Kingdom *see* **Great Britain**
United States
Related headings:
American

United States
Adolescents. Reaction to parental authority. - *Social problems perspectives* 014
Adolescents. Sub-culture formation, search for personal values, religious behaviour, delinquency - *Social problems perspectives* 014
Educational change. Role of administrators 010
Educational planning. Administrative aspects 010
Elementary schools. Role & duties
for trainee principals
Government. Federal programmes.
research 098
High school students. Student involvement in extra-curricular activities. Influence of school size & degree of teacher undermanning 056
Higher education. Prediction of organisation. Prediction models: College organisation based on transfer of

Universities. United States
Authority of administrators. Influence of university growth, demands for accountability & administrative specialisation 057
University growth. Influence on authority of administrators. Universities. United States 057
University of Bradford. *Sample population*
Engineering colleges. Sandwich course students. Deteriorating attitudes to industrial training. Influence of non-skilled routine work in firms & lack of contact with work superiors 035

University of California, *Berkeley*
Research workers. Attitudes to professional status - *Analysis of questionnaire* 012
Research workers. Educational background, Job conditions & Duties - *Analysis of questionnaire* 012
University organisation. Model for predicting organisation of higher education. United States
University organisation based on research *compared with* College organisation based on transfer of learning 088

University students
Work values. Experimental trial of Super's Work Values Inventory - *Sample population: Scottish student groups on vocational & heterogeneous courses* 017
Urban adolescents. Great Britain
Increased illegitimacy rate *related to* Cultural change & Affluence 001
Urban communities
Neighbourhoods. Social characteristics of neighbourhoods *related to* Planning 030
Urban comunities. United States
Social improvement programmes. Effectiveness of schools. control of schools 031
mple population
Innovation research. Development & validation of indices. Indices tested: Number of innovations adopted, Extent of adoption & Earliness of adoption - *Sample population: Urban elementary schools - Study region: Western Canada* 048
Urban origins. Influence on enrolment. Higher education. Russia

Figure 2. An Example of Indexing in Depth

change because through MARC the central system is now offering Library of Congress class numbers and Library of Congress subject headings and a fair number of British university libraries will find this information very useful. This, however, is sheer illusion. What has happened is that a government department has now accepted the responsibility for providing this information as a part of the MARC package. It really has nothing to do with MARC. Now that the principle is established, we can point out that a far larger number of libraries use UDC, and providing we describe it as part of the MARC Project, the two or three thousand pounds needed might be forthcoming. MARC is a trigger mechanism in more ways than one.

Since the real changes that we are witnessing in library systems are due to the impact of automation rather than MARC, there is good reason to pause before designing that totally automated library system. The wind of change is or should be blowing harder at the center than at the circumference. This is particularly true in England where the central bibliographic organization can be entirely outward looking. We are not part of a great national library with enormous responsibilities to that library. In fact the link between BNB and the Library of Congress is much stronger than our link with the British Museum. I was forcibly reminded of this the other day when we had to consider the position of the shared cataloging operation in BNB's new 1971 system design. We are also, for example, already considering profound changes in our interlending and union catalog structure which will have a significant effect in library systems. BNB is changing the whole basis of its card service to respond to standard book number ordering (numbers may be picked up from the book, from publishers advertisements or from the MARC-produced "Books of the Month and Books to Come" provided by Whitaker & Sons). BNB is considering automatically transferring the information on a MARC file which tells us how many tracings are needed for a complete set of entries across to the standard book number file so that libraries can order a set or multiple sets of cards. And BNB has been producing the card service masters from MARC for well over a year already.

These projects are not all associated with MARC but they are a direct result of the British MARC Project. A small data processing team of four people has been placed in a position where it can make the maximum impact on the whole structure of our centralized bibliographic services. We are at the beginning of a complete realignment of the balance between central services and local systems. I must stress that this is just the beginning; the time scale is impressive, perhaps alarming. We appointed our first programmer scarcely eighteen months ago. I do not know what will happen at the center in the next five years, but it is certain to profoundly affect the local use of computers in libraries. This point can be illustrated

rather aptly by describing a new MARC base project which we hope to be testing in libraries in a very few weeks time.

"BOOKS IN ENGLISH"—A MARC DATA BASE SERVICE

One of the major problems associated with producing a national bibliography is the sheer weight of material that has to be processed. British output is now over 30,000 items per year. Because of this, printed bibliographies are getting later and later. It now takes at least six months to produce a BNB annual volume and the multi-annual volume takes two or three years. This situation can be improved by using a computer, but once the final negative is produced on the Photon or Digiset the whole process reverts to a traditional printing and binding cycle which can go on for weeks or months. The input and output technology is sadly out of phase. What we need is a fast method of transferring information from a magnetic tape to a visual form that can be quickly and cheaply reproduced. It is easy enough to get from magnetic tape to microfilm directly, and this looks like a possible solution for in-house authority files and local catalogs, but microfilm is not really convenient to handle and we cannot make 5,000 copies for distribution.

We have instead investigated the possibilities of Photo-Chromic Micro Image (PCMI). PCMI is a microfilm of a microfilm which will hold about 3,000 pages of print on a single transparency. A transparency can be produced very quickly and reproduced very cheaply. It has very definite possibilities. We were anxious to test these possibilities and at the same time to test in a real situation the compatibility of BNB and LC records. I know these records are totally machine and program compatible. Our programs run just as well on the LC file as they do on the BNB file. But can we automatically create highly organized catalogs from the joint file, or will the unpredictable nature of the headings on the Library of Congress file create bibliographic chaos? I sincerely hope not because BNB is about to experiment with a new international bibliography compiled directly from the LC and BNB MARC data bases. The transposition from magnetic tape to PCMI will be direct, without editorial intervention. This bibliography has two special features: 1) every issue—and there are six per year—is a complete cumulation of everything previously published and 2) it is only produced in PCMI form. The annual volume, if that is the right word, will consist of two 6 by 4 inch transparencies. These will hold 50,000 or so entries arranged in a classified sequence and an alphabetic sequence with entries under all tracings. It should not take more than a week or two to produce and distribute. This bibliography which we will call "Books in English" will obviously provide a magnificent visual index to the Library of Congress and BNB MARC data bases. It may also replace

a large number of library catalogs. Why should a large system that would more or less expect to obtain any English-language book that was requested start building these enormous catalogs that duplicate at a frightening cost most of the material in the national bibliography? Maybe it would be better to use an English-language bibliography as a catalog. This idea is scarcely new but it can only be applied if the bibliography contains a classified arrangement, or at least a classification number so that a shelf location is available. Also the cumulating cycle must be much better than has been achieved in the past; with PCMI it can be very good indeed.

This one new service could completely change the design and objectives of a local library automation system. Instead of worrying about updating vast catalogs in the machine system, the librarian may decide to set up an on-line interrogation unit next to a PCMI reader. This would simply contain a string of Library of Congress card numbers and SBN numbers plus locations. The catalog user would identify the book he was interested in and interrogate the on-line files by punching in the LC or SBN number. I do not have the faintest idea what total cataloging costs, including filing, storage space, etc., are like in the United States, but the possibilities of perhaps ceasing the cataloging of 90 percent or so of English-language intake and still providing a good catalog are surely worth investigating.

MARC OBJECTIVES

By carrying through these projects BNB is, by force of circumstances, adopting a general policy of achieving economic goals at the center. Money is tight and we have to exploit MARC as we go along. However in 1971 when the heavy cost of producing MARC records shifts almost entirely from the MARC Project to BNB, then increasing attention and money can be concentrated on the utilization of the exchange service by libraries. England, like the United States, has a very large number of libraries with access to computer systems, but unlike the United States there are at most half a dozen (excluding special and industrial libraries) with sufficient experience to develop automated systems for themselves. With such a small group it is not difficult to coordinate development and our British government policy has been to give special support to three or four systems to enable them to pay special attention to making sure that their development work is as generalized as possible in terms of programming and machines. There are particular machine problems as British computers are totally incompatible with IBM. This means that almost everything will have to be done twice, but at least we can try to insure that it will not be done half a dozen times.

Because of the scarcity of resources and experience, a good deal of the IBM development work will probably be done by the BNB MARC

Project team. This makes sense because we have already gotten a tremendous amount of experience in using MARC to construct catalogs and indexes, plus a sophisticated range of fully operational programs. It would be an enormous waste of time and money if these programs were not made generally available and so a conversion program has been initiated which switches a record in the communications format back into our own internal working format. We have also just finished testing a generalized input program for IBM libraries. Local input is a necessary and rather complex feature of any local system. In my experience libraries usually learn to handle MARC records long before they are able to create compatible records. The use of the generalized input program will accelerate progress and encourage a high level of MARC compatibility and standardization.

MARC STANDARDS

This brief survey of MARC developments at BNB has ended with the word "standardization." Standardization lies at the heart of the MARC Project. It is the essential grammar of communications systems. Most of the ability to construct and manipulate bibliographic data in machine systems has or can be learned by those of us working in this field. Have we however given enough attention to the application and development of standards? Perhaps it all depends on what we mean by standards. I divide them into international standards, national standards and local standards. The MARC record produced by BNB and the Library of Congress contains excellent examples of all three. At the international level we have the Dewey classification number. At the national level we have the Anglo-American code (British text). At the local level we have the BNB classification number, the Library of Congress classification number and Library of Congress cataloging. This classification may seem a trifle arbitrary but I think it would be misleading to regard Library of Congress cataloging and classification standards as national standards. The Anglo-American code (American text) is a national standard but it has not been adopted by MARC. In fact there is the worst possible situation—total uncertainty as to the heading we can expect on a MARC record.

As far as the Library of Congress classification and subject headings are concerned, I would be happy to see these systems elevated to the level of national or even international standards if only they could be developed quite independently from the Library of Congress collection. Generally speaking, it is possible to handle the list of subject headings as a standard. The authority exists and the rules for creating new headings are reasonably explicit. It is however quite impossible to treat the Library of Congress classification schedules

as a standard. In attempting to do so we have had to devise a set of rules which separates the classification structure from the local system elements. They are unfortunately closely interwoven, but they can be separated. The classification scheme has such obvious merits that one of the greatest services the Library of Congress could render is the preparation of a public standard version of its own schedules and the use of it in MARC.

Quite clearly MARC has not yet acquired the essential grammar of communications systems, even between two countries with a common language and very closely aligned bibliographic practices. What then are the chances of achieving standardization and compatibility in 1971 when European MARC systems will be developed? An international MARC system should be an exchange of bibliographic data compiled according to a consistent set of principles between national agencies. We need a standard communications format, a standard set of cataloging principles and a standard means for signaling subject content. We must, at least during the first international expansion of MARC, use existing standards. Fortunately these standards are available.

THE COMMUNICATIONS STANDARD

The existing communications format used by Britain and America can be regarded as a satisfactory international standard. Some of its features make it rather more satisfactory for use at an international level than at a national level. In fact as part of a general MARC II evaluation program we are considering whether the national and international communications standards should be logically distinct. We want to receive tapes from the Library of Congress and European MARC agencies in a standard form, and we are prepared to return tapes in the same standard form, but why should it be assumed that communication between the Library of Congress and the University of Illinois, and between BNB and say, the Bodleian Library should be in exactly the same form again. As it happens we already use a special national communications format for the libraries in the British network that are using British computers. This format is much more convenient to process than the standard communications format. When we go into Europe we should only be concerned with communications at the international level.

THE CATALOGING STANDARD

National cataloging standards are embodied in detailed codes of rules. An international cataloging standard must be a set of principles. It is vitally important to the future of MARC that a common set of principles form the basis of MARC cataloging. These principles exist; they were drawn up at the Library of Congress by Seymour Lubetzky. They were presented as a "statement of principles" adopted by the International Conference on Cataloguing Principles,

Paris, October 1961. They form the basis of the Anglo-American (AA) rules, and the British text faithfully follows these principles; the American text does not. The reason is given in the preface: "It is regrettable that, because of the great size of many American card catalogs, it was necessary for the Catalog Code Revision Committee to agree to the suggestions of the Association of Research Libraries that certain incompatible American practices be continued in the present rules."[3] "Regrettable" in the national context perhaps, although "disastrous" would be a better word in the context of MARC and an international MARC network. It is perhaps fortunate that the AA code has not yet been accepted by the Library of Congress and that a change to the basic set of principles is still a practical proposition. For the international MARC network the Paris Principles are surely the only possible cataloging standard.

THE SUBJECT IDENTIFICATION STANDARD

It may seem an impossible task to reach agreement on an international standard for subject representation on a MARC record. In fact as long as we have a clear understanding of the purpose of the standard it should be much easier to achieve than a cataloging standard. National standards are not involved at all. What we need is a common means of conveying subject information in MARC. From a logical standpoint the information carrying system needs only be fairly precise in its identification of complex subjects. From a practical point of view this system should be familiar and in widespread international use. It should also be an adequate working tool for subject retrieval in a machine system. Only one existing scheme meets these requirements—the Universal Decimal Classification (UDC). UDC is capable of considerable precision in subject identification and is widely used in America, Great Britain and continental Europe. It has retrieval capabilities which are far superior to DC or LC. A sound economic case could almost certainly be made for adding UDC as a national class number together with LC and DC on both British and American records. Many libraries in both countries use UDC. Many other systems would use the UDC data for selective dissemination of information (SDI) services. In the context of an international MARC network UDC would be invaluable because the UDC number contains enough information to permit a very high level of conversion to local subject systems. In Great Britain we have neither the money nor the inclination to handle MARC as if it were a machine-readable version of the shared cataloging program. If Paris or Rome or Stockholm sends us national MARC data we would only expect to augment this data by machine intervention before adding it to our international MARC data base. This is perhaps the fundamental difference between the British and American approach to MARC and the main reason why MARC standards seem to us to be so

vitally important. BNB will not recatalog or classify any material coming over the international network but it will be prepared to convert an international standard subject representation to national system equivalents. Conversion systems depend absolutely on the built-in precision of the starter scheme. For practical purposes this precision is available in UDC.

REFERENCES

1. U.S. Library of Congress. Information Systems Office. *MARC Pilot Project: Final Report.* Prepared by Henriette D. Avram. Washington, D.C., Library of Congress, 1968, p. 1.

2. Avram, Henriette D., and Curran, Ann T. "The Identification of Data Elements in Bibliographic Records." (n.p.), 1967.

3. *Anglo-American Cataloging Rules.* Prepared by the American Library Association, the Library of Congress, the Library Association, and the Canadian Library Association. North American Text. Chicago, ALA, 1967, p. vi.

Josephine S. Pulsifer
Chief, Technical Services and Development
Washington State Library
Olympia, Washington

MARC BOOK CATALOG PRODUCTION IN WASHINGTON STATE

THE LIBRARY PICTURE IN WASHINGTON STATE

PROFILE OF LIBRARIES

The current inventory of libraries in Washington State compiled by the Washington State Library shows the following profile:

College and university libraries		18
Community colleges		22
City, town, county, and regional public libraries serving a population of:		
Over 100,000	12	
25,000-100,000	8	
5,000-25,000	18	
Under 5,000	45	
Total		83
Business and industry		22
County law libraries		39
Historical societies and museums		5
Medical libraries		8
Other special libraries		3
U.S. government libraries		18
Washington State government libraries		4
		222

The figure 222 represents a minimum count since only parent institutions or organizations are included. Branches of main libraries or separately administered libraries in different departments or locations are not counted; however, Air Force, Army, Navy, and veterans administration installations are counted separately. Of the college and university libraries, six are state supported. Included in this figure is the new Evergreen State College scheduled to open in Olympia in 1971. All of the community colleges have been under the jurisdiction of the state since 1967.

THE STATE LIBRARY

The State Library Commission, appointed by the governor, is the policy-making body for the State Library. The state librarian is a member of the governor's cabinet.

The State Library is organized into four divisions: management services, readers services, technical services and development, and library development, all reporting to the state librarian. An associate librarian for planning and research was appointed in July 1968.

Except for the law library and archives services, which are the mandate of separate agencies, the State Library has the responsibility for all library services to the legislature and to all agencies, institutions, and other organizations of state government. It also is charged with the development of library service and interlibrary cooperation throughout the state.

The latest addition to its services is a training and audiovisual services program established for the purpose of providing non-print reference service and serving as an instructional resources center for the state government. Within the latter responsibility, a pilot project has been undertaken to provide technical and material support to the department of personnel's training division, which serves 30,000 state employees.

LIBRARY DEVELOPMENT

A state-wide library development plan was formulated by the Washington Library Association (WLA) as early as 1934.[1] The State Library Commission and WLA updated and made more specific the public library portion of the state-wide plan following a study by Charles Bowerman in 1948, which recommended division of the state into twelve regions, each with population and tax bases large enough to allow it to maintain adequate library service.[2] Funds from the Library Services and Construction Act (LSCA) and its predecessor, the Library Services Act, have since supported three multi-county regional library "demonstrations," all of which have resulted in the assumption of support of library service by vote of the people.

INTERLIBRARY COOPERATION

The outstanding example of interlibrary cooperation in the area is the Pacific Northwest Bibliographic Center (PNBC) founded in 1940 to serve the states of Idaho, Montana, Oregon, Washington, and the province of British Columbia—and now Alaska. While the center's activities have been somewhat curtailed from the original program, the center serves as a major "switching center" for interlibrary loan among members. Washington libraries are by far its heaviest users. A recent study by Lura Currier analyzes PNBC's present status and makes recommendations for its future role, funding and operation.[3]

Further highlighting the cooperation that exists at the policy-making level in the state, the Washington Library Association's planning committee for state-wide programs has long acted as an advisory council to the State Library Commission in the interest of developing good library service in the state. This committee has recently been combined with the Title III (LSCA) advisory council to form the state-wide library development council, charged with advising both WLA and the State Library Commission on library development and interlibrary cooperation programs in the state.

In addition, the Washington Higher Education Library Committee (WHELCOM), composed of the librarians of the state-supported colleges and universities and the state librarian meets regularly to consider matters of common concern. The willingness of the large university and public libraries to cooperate with the State Library in the development of state-wide library services is long standing.

Shared Resources– At the present time, all interlibrary loan requests from public libraries in Washington are channeled through the State Library which purchases nonfiction materials as back-up for public library collections as well as for its own services to state agencies.

Interlibrary loan requests not filled by the State Library are forwarded to PNBC whenever the requesting library is a member. Requests for interlibrary loan from non-public library members usually go directly to PNBC. Washington State members of PNBC number eighty-seven, but only nine of those libraries contribute their holdings. The University of Washington holdings are not incorporated into the union catalog, but its catalog is available for consultation.

Agreements for reciprocal services are in effect between Spokane Public Library and Spokane County Library, Bellingham Public Library and Whatcom County Library, Port Angeles Public Library and Clallam County Library, Seattle Public Library and King County Library System, Tacoma Public Library and Pierce County Library. In each case, a large municipal library has opened its doors, for only a token fee, to give the same service it provides for its own borrowers to patrons of the contracting library.

A common library card is issued to borrowers of the tri-cities area for use in the Pasco Public Library, the Richland Public Library, and the Mid-Columbia Regional Library in Kennewick, serving both Benton and Franklin counties.

Interconnection of the state-supported universities, four-year colleges, community colleges, and the State Library on the SCAN (State Controlled Area Network) telephone system allows inexpensive query among those libraries, which is utilized to some extent informally. SCAN institutions may call off-network libraries, but off-network libraries have no access to the network at the present

time, though there is a possibility that libraries and other special groups may be admitted as members at a later date. Special courier service has been established between the University of Washington and Western State College in Bellingham. The Boeing Company contributes funds for the purchase of aviation materials by the Seattle Public Library and also supports a staff member at the University of Washington.

A study of *Library Service Across the Border: Idaho, Oregon, Washington* was conducted during the summer of 1969 by Grace Stevenson.[4] The study recommends a four-county demonstration made up of two counties each from Oregon and Washington, experiments in the use of a common library card and numerous other cooperative projects which would serve users on either side of the common borders.

Shared Technical Services– There are no processing centers in the state which serve multiple independent libraries. Regional public library systems perform centralized processing services for their branches, as do the large municipal libraries and school systems. Recently Anacortes Public Library agreed to assume processing services for LaConner. Battelle Northwest Technical Information at Richland performs technical processing for the Graduate Center for Research, a graduate school sponsored jointly by the University of Washington, Washington State University and the University of Oregon.

Since 1965, the State Library has performed all technical services for the five-county Timberland Regional Library (TRL), formerly Timberland Library Demonstration. Prior to that date all catalogs of the member libraries were merged into a union book catalog. A current book catalog is published bimonthly, cumulative continuously throughout the year. A complete reissue of the TRL catalog through 1969 is in progress.

Since 1966, the State Library has also published the book catalog for North Central Regional Libraries (NCRL) and now provides catalog data from LC proofslips or original cataloging for all new NCRL titles, but neither orders nor physically processes their materials. Beginning in January 1970, the Timberland and North Central book catalogs have been combined into a single catalog which shows the libraries holding each title. The experimental MARC book catalog produced for these two library systems and the King County library system is described later in this paper.

WASHINGTON NETWORK PLANNING

THE CONCEPT

The previously described activities occurred quite independently of the impetus toward interlibrary cooperation generated by Title III

of LSCA. Specific planning for an interlibrary cooperation program within the state under LSCA began as soon as Title III was introduced in Congress, and the advisory council was appointed immediately upon passage of the bill. Early in 1967 the council requested that the State Library Commission engage Joseph Becker and Robert Hayes to develop a state plan.[5] Basic objectives were agreed upon between the consultants and the council, and the plan was presented orally to the council in July 1967, and approved in concept by the council and the commission. The concept was presented to the profession at a series of fourteen meetings around the state, and to the public at several governor's conferences on libraries between September 1967 and May 1968.

LSCA Title III was merely the catalyst; the real impetus for a network stems from the increasing demands placed upon libraries and the resulting need to use available resources better. In the words of the Becker and Hayes plan:

The intent of the program is three-fold:

1) To promote the increased sharing of resources by libraries, particularly of different kinds and with different area jurisdictions.
2) To use modern technology in an appropriate, economic manner and by doing so, to facilitate the sharing of resources.
3) To expand the availability of library materials to every resident of the State.[6]

THE BECKER AND HAYES PLAN

The library network as proposed by Becker and Hayes is constructed of groups of libraries cooperating voluntarily, independent of their administrative base, on the basis of geographic area. One dominant library in each group is identified as the group center. Directories of the holdings of all member libraries in the area group allow maximum sharing of resources within the group before going outside the group. Directories of the other area groups in the area center make it possible to identify possible sources of interlibrary loan if materials are not available within the local area. A catalog production center at the State Library produces the directories for the various area groups, and may perform other centralized cataloging or processing services. A central switching center provides switching and referral services for requests that cannot be satisfied at the area group level and which must therefore be referred to major state and national resources. Communication equipment differs at each level depending on the kind and amount of use demanded.

Groupings of certain libraries are also proposed based on subject specialty or specialized constituency served. For example, a library with an important collection in medicine or a community college library would join with other public, academic, special, and school

libraries to form a geographic area group, but the medical library might also join with other medical libraries throughout the state in a medical specialty group and the community college library with other community college libraries in a community college specialty group.

NETWORK REQUIREMENTS

An efficient network depends, among other requirements, on 1) knowledge of what materials, expertise, etc., are available and where they are located; 2) agreements as to what materials may be borrowed within the network; 3) fast, efficient internal procedures for filling requests; and 4) common formats, procedures, and languages.[7]

The communications aspect of networks is often overemphasized at the expense of the basic information of *what* and *where*. Broadcasting a request for an item with no knowledge of whether it is available within the network, or where a copy of the item may be located is a very wasteful activity. This is not to say that an improved communications system among libraries should not be undertaken until there is a capability for determining what is where, but it seems that the directory function should have at least equal priority with the upgrading of communication facilities. Ordinary telephone service (while admittedly lacking the virtue of producing a hard copy record) and mail service can be very effective when combined with prior knowledge of holdings and speedy processing of requests *within* the libraries on either end.

For the directory to function effectively, the bibliographic items should be described and analyzed according to a standard set of rules, against a common authority file and data base, with a common degree of completeness. Further, the machine-readable record must be structured in a standard format with standard content designators for each type of material.[7] The MARC format provides the standard structure. Standard content designators for monographs, serials, maps, and motion pictures and filmstrips have now been drawn up. The MARC Distribution Service provides bibliographic records in MARC format for English-language monographs currently cataloged by the Library of Congress. Future extension of the Distribution Service to other languages is dependent upon the extent of support evidenced by the library community and the consequent ability of LC to secure funds.

The RECON Pilot Project will provide the same standard MARC records for LC English-language monographs back through 1968 or earlier. A future conversion program depends upon the results of the pilot project and, again in large measure, on the support of the library community.

The Washington State Library, as the agency responsible for developing and administering the network, is firmly committed to the use of the MARC format and MARC records as the basis for network

services. Currently acquired materials which are not within the province of the MARC Distribution Service will be locally input to be compatible with MARC. Furthermore, the State Library considers it uneconomical and wasteful to consider conversion of retrospective records before the RECON Pilot is complete and before a determination is made of subsequent conversion activities by the Library of Congress, unless the circumstances are unusual. The library intends to make maximum use of and cooperate in all possible ways with a future RECON program.

SYSTEMS DEVELOPMENT FOR THE NETWORK

MARC PILOT PROJECT

The system now being designed to provide 1) current directories of holdings of network member libraries, and 2) current cataloging support services is a direct descendant of the State Library's MARC Pilot participation. As noted earlier, the State Library was already producing book catalogs for Timberland and North Central areas. King County library system had been producing book catalogs for its branches since 1951. A book catalog system was therefore a logical experimental project for Washington State. Since the same catalog data was required for shelflists and book preparation materials, the State Library undertook to provide these materials, as required, for participating libraries. Complete sets of cards were printed for the State Library catalog. While the cards and labels were printed and distributed weekly, the combined three-system book catalog was produced as a one-time output, covering titles acquired during 1967. No effort was made to input titles that did not appear on the tape, and since it was a pilot project, all of the three individual catalogs were run as parallel systems throughout the experimental period.

The catalog format chosen was that of a register with indexes. This type of catalog had already been produced commercially using a photographic reproduction of a printed catalog card as the register entry. To the best of our knowledge, Washington State was the first to produce both the register and index from a machine-readable record.

The four sections of the catalog may be described as follows:

Section 1: A register, consisting of full LC entries arranged by a register number assigned in the order in which a title came into the system. With this arrangement, the full entry catalog could be constantly added to but would never require resequencing and reprinting to accommodate new titles.

Sections 2, 3, 4: Author, title, and subject indexes or finding lists. These briefer entries gave enough information to identify the book, locate it on the shelf, or request it on interlibrary loan.

Each index entry gives the register number, so that the user can refer to the full bibliographic information as he wishes. All LC tracings were used as access points (with the exception of series added entries, which were omitted only because programming time was limited). Additional subject headings, and author and subject cross references were added as needed. Each entry in the index sections showed mnemonic codes for libraries holding that title.

The experimental catalog proved the utility of centrally produced and distributed MARC tapes for the production of book catalogs, although no attempt at any real resolution of the filing problem was made. The IBM Sort 7 program was used to operate on a sort field created automatically for each entry. The specifications for the sort fields are shown in Table I. In addition, a table look-up was used to suppress initial articles in English for filing purposes.

Table I. Sort Field Specifications for Washington State Library MARC Pilot Book Catalog

	Author field	Title field	Subject field	Date field
Author catalog	40	16		4
Title catalog	16	40		4
Subject catalog	15	5	70	

The specifications for the experimental catalog purposely precluded any "editing" of the MARC record that could not be done automatically by programming. This was done to take full advantage of the centrally prepared catalog record and to demonstrate the greater economy of using the standard record. All three public library systems and the State Library accepted the LC author entries, subject headings, and an automatically shortened Dewey Decimal Classification. Dewey numbers were added if lacking in the LC record.

Since automatic acceptance of LC cataloging obviated any need for "proofing" a printout of the MARC record before printing the cards and pockets, all cataloging support services output could be produced as soon as a matching MARC record was found. Books could thus be processed and put into circulation with minimum delay following receipt.

The State Library MARC system was programmed in Autocoder for the IBM 1401 computer with 14K units of core storage then available at the State Data Processing Service Center. Later in the project IBM 360 models 30 and 40 were made available to us in

another agency. The programs were modified for running in emulation mode on those computers in order to take advantage of a 60-character print chain and a lower hourly charge.

The card print program provided to MARC participants was utilized, with modifications, to print either single cards or card sets and the entries in the register section of the catalog. For the latter, extension cards had to be made continuous with the first card, and "white space" normally left between tracings and body of the entry had to be compressed.

BASIC BIBLIOGRAPHIC SYSTEM FOR NETWORK

The basic bibliographic system to be designed is a generalized set of programs to allow maximum flexibility in inputting, processing, and outputting bibliographic data and associated information. Within the broad framework, three subsystems have been identified for implementation: 1) directory subsystem, 2) catalog support services subsystem, and 3) in-process information subsystem.

The development will be a joint effort of the State Library, the Evergreen State College, and King County library system. In addition to these three libraries, the first group of libraries included in the project is expected to be the two public library systems for which the State Library's Technical Services and Development Division already performs cataloging services and probably two additional systems to which service may be extended during the coming year. A community college, public school system, and a special library should be added to the group at a slightly later date, in order to test the system with all types of libraries.

Complete technical services are expected to be provided only for the State Library and for the Timberland Regional Library. The manner and extent to which the Evergreen State College and the State Library will work together in matters of acquisitions and physical preparation is not fully determined at this date. As presently foreseen, network services will be confined to directory and catalog support services. The latter will be optional, and by contract. Member libraries not desiring this service will be required to submit their catalog data for entry into the directory. It should be understood that we are chiefly concerned with currently acquired library materials at this time, not with retrospective materials, although an opening-day catalog for Evergreen and an urgently needed retrospective catalog for King County are priority requirements of the system.

Directory Subsystem– The directories will provide bibliographic and location (holding library) information for groups of libraries for the purpose of making materials held by libraries in the group more accessible to all borrowers. Whether the group directory is as

acceptable to the public and to the librarians as the catalog of the individual library must be determined by experimentation. Instead of group catalogs combining bibliographic and location information, all of the bibliographic information for all titles held by network members anywhere in the state—or by those libraries covering a major portion of all titles held—could be printed in one catalog. Location information might then be printed as separate listings by area. This latter pattern of catalog organization requires a double look-up operation for location data, but avoids duplication of the entries when the same item is held in more than one area group.

There will also be monitoring of the utility of the register/index catalog, which has yet to be tried in an operational environment in Washington. To avoid different register numbers for the same title in different group directories, the register number must be assigned on a state-wide basis and a single register issued. It is possible that only the indexes are needed in many libraries, and that the register is required only in larger libraries and group centers. Possibly the National Union Catalog (NUC) can substitute for the register as a source of the full entry. That there is interest in the register/index style catalog for a future NUC compounds this problem, since local indexes would then have to refer to the NUC register number which could not readily be known in advance.

Among further questions to be answered are the following: Should serials titles (other than periodicals) be included in the same directory with monographs, leaving holdings data for serials union lists, or should they be separated entirely? Would it be as desirable as generally thought to include phonorecords, motion pictures, and other non-book materials in the same catalog with monographs, or are they best consulted separately? Should adult and juvenile titles be listed together, as current opinion among two of Washington's regional libraries suggests, or is this unsatisfactory? And are there differences in the answers to any or all of these questions depending upon the size and type of library? These are some of the problems Washington State hopes to gain insight into as it begins implementing its directory system.

The production technique of book catalogs will also be the subject of experimentation. The entire catalog may be printed by line printer, photo-reduced, and reproduced on an offset press; all or certain issues may be printed using photocomposition techniques; computer-output-microfilm techniques may be used to create a catalog on microfilm or go through microfilm to a master for offset reproduction. Microfiche is another possible medium for publication.

Depending upon the method and cost of production, an optimum frequency of issue and pattern of cumulation must be determined. Present plans for the initial project call for monthly issues, cumulated continuously up to one year.

Cataloging Support Services– This service will include the distribution of selected MARC records in printed or machine-readable form, including catalog cards, labels, punched circulation charge cards, magnetic tapes, etc. These are provided according to a profile of requirements for each library.

Individual library book catalogs in standard format and using standard catalog data may also be produced as subsets of the directory subsystem. The catalog of current acquisitions of the State Library will in all probability be separately produced for distribution to all state agencies and to libraries throughout the state to facilitate borrowing from the State Library.

In-Process Information System– This subsystem must operate on three levels. Where ordering, receipt, and processing of materials are required, it serves to produce order lists by vendor, and weekly updated status lists showing titles which have been ordered, received, and processed. At each stage, the receipt of a matching MARC record will be indicated. Titles will be deleted from the in-process list upon their appearance in the book catalog. In addition to serving as a status information system, this subsystem will provide management information concerning on-order time for purposes of vendor evaluation, MARC record receipt time in relation to receipt of the book, and processing time.

Where only cataloging support services are required, the in-process subsystem will report entry of the request into the system and shipment of required products out of the system, along with information on availability of MARC data. Titles reported purely for directory purposes will be listed as in-process until they appear in the book catalog update, with the indication, where appropriate, that there is a matching MARC record.

At present, the in-process information system is envisioned chiefly as a control and management tool. At a later date, the system might be expanded to provide information to the individual library on the status of its requests for materials or catalog data, and to the group center on materials in process in its group between updates of the book catalog. Because of the volume of updating, this subsystem is an obvious candidate for later on-line operation.

When future acquisitions and on-line circulation systems are implemented for the State Library, hopefully in common with Evergreen State College after the basic bibliographic system is operational, the in-process information system will become a link between the two.

USE OF MARC RECORDS

MARC AS THE STANDARD RECORD

Some policies which are being followed in the design of the network system are as follows:

MARC records will be used at as early a point in the system as they are available, and with effectively no change except for certain programmable options. MARC records are used without prior prooflisting, with minor exceptions.

MARC records will be accessed only by the LC card number until a standard author/title search code for use with them has been developed for the library community.

Temporary cataloging will be done for materials received in advance of an expected MARC record when there is a request or reserve on the book, or after a predetermined elapsed time, but the temporary record will be superseded by the MARC record when received.

Local cataloging of materials not cataloged by LC will be performed as nearly consistent with LC practice as possible. Where libraries contribute their own local cataloging for directory listing, MARC and/or LC data will be substituted when they can be located; where no LC data is available, the local record will be entered as is with its source identified as a prefix to the identification number.

All records input by the State Library cataloging center will be given content designators effective up to the level of MARC records edited by LC. Future uses of the data base are unknown; we do not wish to reduce the level of content designation until a reduction is jointly agreed to by LC and the library community.

TIMELINESS OF THE MARC RECORD

Timeliness of the MARC record—and prompt utilization of those records that are needed—is of primary importance, so that library materials may be processed and put into circulation in the shortest possible time. In order to have the critical items of catalog data available at an earlier date, Washington State Library urges that cataloging-in-source be reinstituted, that the MARC record be created at that point, prior to availability of the collation data, and that the MARC record be updated for collation information only for LC internal use. If updated records are distributed at all, this might be done at an extra cost to subscribers.

TECHNICAL ASPECTS

The State Library is required to use the facilities of the State Data Processing Service Center. The computer facilities consist of an IBM 360 model 50, under O/S (operating system), operating in an MFT (multi-programming, fixed number of tasks) environment. High speed tapes, 2,314 discs and remote terminal hook-up are available as needed. A "TN" text printing print chain with 120 characters, consisting of upper and lower case alpha and special characters is also available.

The participating libraries are working together on final specifications for the system. Whether the decision is made to have the computer system design and program contracted out or done "in-house," permanent staff of Evergreen, King County, and the State Library will monitor its programs and be thoroughly trained for continued operation of the system.

REFERENCES

1. Washington Library Association. *A Program for Library Development in the State of Washington, December, 1934....* [Seattle, 1934].

2. Bowerman, Charles E. *A Proposed Regional Library Plan for the State of Washington; Report of a Study of Public Library Needs in the State of Washington.* Olympia, Washington State Library, 1950.

3. Currier, Lura G. *Sharing Resources in the Pacific Northwest.* Olympia, Washington State Library, 1969.

4. Stevenson, Grace T. *Library Services Across the Border: Idaho, Oregon, Washington; A Study.* Olympia, Washington State Library, 1969.

5. Becker, Joseph, and Hayes, Robert M. *A Proposed Library Network for Washington State; Working Paper for the Washington State Library.* Olympia, Washington State Library, 1967.

6. *Ibid.*, p. 3.

7. Avram, Henriette D. *Bibliographic and Technical Problems in Implementing a National Library Network.*1970. (In process of publication.)

Frederick H. Ruecking, Jr.
Head, Data Processing Division
Fondren Library
Rice University
Houston, Texas

MARC AT RICE UNIVERSITY

Much of the experimentation associated with the MARC Pilot Project at Rice University has been continued into the MARC Distribution Service. Consequently, this review of activity incorporates results from studies using both formats and data. That these studies have been oriented toward retrieval problems is a product of circumstance.

At the time the pilot project was started, the library was not particularly interested in the production of catalog cards or book catalogs. The primary interest was, and continues to be, the development of aids to the regional exchange of resources. One of the early dreams in library automation was user access to a bibliographic file via an on-line, remote terminal.

The magnitude of a regional base, the required latitude in controlling search input, and the variations of supporting logic suggested that our interests could be served best by studying the characteristics of user-based input and examining alternatives for retrieval. Execution of these studies generated a broader perspective and involved the library in special projects distantly related to intended purposes.

This paper reflects this broadening of perspective. The initial studies were involved with search codes, i.e., compression algorithms designed to yield proper citations from faulty search input. A second series centered upon access methodology, while the third and final series (to be reported) concerned the organization of a library system.

Segments of the studies reported were funded by the National Science Foundation, the Texas State Library, the Regional Information and Communications Exchange, and Rice University.

SEARCH CODE STUDIES

A search code may be defined as an abbreviation which serves to identify a specific record. As used in this paper a search code is defined as the product of an algorithm which utilizes bibliographic

data to produce a limited character-string which can be used to uniquely identify a given item.

Since the algorithm must operate on data base and user-generated input, it must provide for the reduction of user error by offsetting mistakes in spelling and producing more opportunities for matching than generally employed in the current cataloging system.

The code strings described incorporate significant title words and author surnames. Corporate and conference entry codes are appended to the title codes. The complete series of possible codes is produced to yield maximum matching capability. Thus, if a work has an author plus two added entries and a traced title and added title, the total number of codes generated will be six.

Four algorithms were examined, each of which has certain advantages and disadvantages for an operational system. These four algorithms have been called the consonant-string code, the three-to-one code, the phonemic code, and the bucket code.

CONSONANT-STRING CODE

Briefly, concerning this code, the compression function is based upon the knowledge that English-language words frequently have initial and terminal consonant strings which tend to identify the word itself (assuming the deletion of prefixes and suffixes).

The test performed at Rice University on a large sample indicated acceptable accuracy without excessive false drop. The primary difficulty was the English-language bias of the algorithm suggesting the possible need for language-oriented algorithms for code generation. The complexity of the generating algorithm, the possible need for multiple algorithms, and the amount of processor time indicated suggested that additional codes should be examined. One of the high-accuracy codes found was a relatively straightforward reduction code based upon known character frequency.

THREE-TO-ONE CODE

There are two variations of this system, each of which appears usable. The first variant requires the use of the first character word as the leading character of the code string. The remaining characters of the word are divided into sequential blocks of three characters each. Using established frequency tables for the language, the least frequent character of each group of three is added to the code string. This simplistic algorithm sharply reduces the total length of the generated code, as compared with the consonant-string code which is always four characters long, unless the word is more than ten characters long.

The second variation utilizes the same frequency tables and uses the least frequent characters of a three-character group. However, this variation does not extract the initial character unless it happens

to be the least frequent. This variation produces a shorter code string than the first variation, and tends to distribute generated codes more evenly through the consonants in the alphabet.

THE PHONEMIC CODE

The phonemic code uses the basic method of the reduction code just detailed. However, the actual characters may or may not be used depending upon its definable phonemic value. For example, the "c" and "h" of the character string "sch" are not separable in English. The character pair is most nearly the "k," consequently, for coding purposes a word such as "school" would become "skul" when the vowel form "oo" is altered.

The algorithm for this construction is much more complex than the three-to-one reduction code, and necessitates more processor time. It does, however, provide some benefit in overcoming errors in spelling and in spreading out generated codes because of the higher number of phonemic symbols available.

Table 1 contains fourteen words and the codes produced by the variations of the systems described. The phonemic form is not exact, the purpose is rather to illustrate the types of character strings generated.

At this point we are unable to suggest which may be more advantageous in operational use. The three-to-one offers the most distinctive codes, but the phonemic codes provide greater dispersion of codes since there are more possible phonemic symbols available. A code which is not based on sequential word order is the bucket code.

THE BUCKET CODE

The principle of the bucket code is the utilization of the title and author stream as a character string without intervening blanks or punctuation. Data are extracted from the title and author with non-significant words discarded, a count is made of the number of characters in the string and blanks are used if the total is less than 128.

An exponential algorithm is used to pick the characters for the code, using "2" as the base value. Therefore characters 1, 2, 4, 8, 16, 32, 64, 128 are extracted to form the code word (eight characters). Carter and Bonk's *Building Library Collections* contains thirty-two characters and produces the code "BULGCR--."

The bucket code is not expected to produce completely unique codes for every item. It is expected to distribute the codes sufficiently which will tend to limit the frequency with which these codes are duplicated.

Our experience in code specification and testing suggests that there may not be a universal in search code formation. The utility of

Table 1. Word Code Comparison

Word	Consonant String	Three to One		Phonemic Form	Variations	
		Variation 1	Variation 2		1	2
CHRIST	CHRI	CHS	HS	KRAIST	KRS	KS
CHRISTENDOM	CHRI	CHSDM	HSDM	KRISTENDOM	KRNM	KSDM
CHRISTENING	CHRI	CHSNG	HSNG	KRISTNIŊ	KRNŊ	KNŊ
CHRISTIAN	CHRI	CHSN	HSN	KRISTYUN	KRYN	KYŊ
CHRISTIANITY	CHRT	CHSNY	HSNY	KRISTYANITI	KRYNI	KYNI
COLLECT	COCT	CLC	LLT	KOLEKT	KLK	KK
COLLECTS	COCT	CLCS	LLS	KOLEKTS	KLK	KKS
COLLECTING	COCT	CLCG	LLNG	KOLEKTIŊ	KLKŊ	KKŊ
COLLECTION	COCT	CLCN	LLON	KOLEKŠUN	KLKN	KKN
CONNECT	COCT	CNC	NNT	KONEKT	KNK	KK
CONNECTION	COCT	CNCN	NNON	KONEKŠUN	KNKN	KKN
CORRECT	COCT	CRC	RRT	KOREKT	KRK	KK
CORRECTING	COCT	CRCG	RRNG	KOREKTIŊ	KRKŊ	KKŊ

any code to be derived from user input may be dependent upon factors which are not under complete control of the library. Such factors as user sophistication, careless keyboarding, the ratio of English to non-English material in the collection, the subjective bias of the collection (if present), and collection size can materially affect the value of any one, or all, of the search codes which can be devised.

ACCESS METHODOLOGY

The most interesting problem and the most difficult to resolve is the formation of a file system which encompasses random needs. Interfiling, referencing, and deleting information from a large file necessitate an access technique which is the foundation for file organization. Three methods have been examined experimentally at Rice University, all of which have particular strengths and weaknesses. The important fact to be recognized is the use of the text itself to provide directional assistance to the record location.

THE HASHED FILE

Hashing is intended to evenly distribute textual data throughout a file which has specific, definable limits. A convenient method of securing such a distribution is the use of a prime divisor. For example, given a file which may contain numeric data ranging from one to 1,000 and will contain 100 different records at one time, the best prime to use is 101; therefore, the file will be defined as containing 101 records.

The actual numeric value of the text is divided by 101. The remainder resulting from that division will be between zero and 100. The actual file location of the record is obtained by adding one to that remainder.

There are certain characteristics which can be used to determine whether or not a hashed file is appropriate: 1) the file must be fairly stable without a great growth rate, 2) the lowest and highest text values must be known, 3) the size of the file must be definable, 4) the frequency of access must be high, and 5) the non-use of allocated storage space must be acceptable and economically reasonable.

The utility of a text-based file of this type is potentially enormous. At Rice University we hope to utilize this basic idea in our on-line circulation system. There are three basic on-line files—current loans, holds, and borrower. There are several off-line files in the form of printed lists and indexes, the most important being the alphabetic borrower index.

The current loan file has a maximum of about 40,000 records, the hold file 8,000, and the borrower file about 8,000. Text limits for the current loan and hold files are AA through ZZ and 000000000 to 999999999 for the borrower file. The current loan and hold file

records are quite short, requiring some sixty-four characters for each one; therefore, to secure more suitable use of storage space these records will be "blocked," 120 records to the block. We will, therefore, provide 337 record blocks since 337 is the nearest prime of 333. Similarly, we will use sixty-seven for the hold file.

Call numbers of individual transactions will be translated into numbers and divided by the appropriate prime. The remainder plus one will be the block in which the record will be found. An internal core scan will be used to ascertain the appropriate position of the record in the block.

The borrower file has longer records and 200 "blocks" will be needed giving an actual size of 211 blocks based on the prime. The social security number segment of the borrower number will be used to compute the location of borrower records.

Since all files have comparable record locations, providing file interplay is quite straightforward. For example, the following procedure stream on a return transaction would be standard.

1. compute record location
2. get record
3. if not overdue go to 9
4. if paid go to 9
5. compute fine
6. compute borrower location
7. get borrower record
8. prepare bill
9. if not held go to 20
10. compute hold location
11. get hold record
12. compute borrower location
13. get borrower record
14. prepare notice
15. delete hold record
16. subtract 1 from hold indicator
17. alter borrower number to hold status in loan record
18. notify attendant to hold and refile loan record
19. go to 21
20. delete loan record
21. exit

There are several additions which would be useful and several alternatives, but the routine series described does provide some idea of the automatic nature of the file interplay. The time frame for the twenty-one steps is machine dependent—but will almost certainly be I-O (input-output) bound. A procedural rearrangement could provide some increase in through-put speed by overlapping the I-O statements.

It is possible to provide considerable variability in hashed file

techniques by using a monitor which examines the use of allocated space on a periodic basis. Such a monitor could scan the file counting unused portions, then through a regenerating system reallocate the space and use a different prime to reduce or enlarge the needed space.

THE STRUCTURAL TREE

The alphabetic nature of bibliographic data suggests that a structural tree would be useful in manipulating bibliographic data. The structural tree is a system which operates in alphabetic sequence using a high-low-equal test at decision nodes. The start-point node can be computable or addressable from the text supplied, but the record containing the matching text can be located only by stepping through successive decision nodes.

Of primary importance in establishing a structured tree is establishing the initial sixty-four decision nodes. An error or miscalculation in the initialization will produce a distorted structure at a later date and one which will be grossly inefficient.

Assume a file of 4,000 names: to construct the initial decision nodes it is necessary to determine the distinctive character-strings of the names at the sixty-fourth interval (i.e., the sixty-fourth name, the one hundred and twenty-eighth, the one hundred and ninety-second, etc.). Once these points are known, the file can be initialized.

Each file record in addition to its text will also contain three directional addresses: 1) the location of alphabetic text of lesser value, 2) the location of text of greater value, and 3) the location of the last decision node. To interfile a record, the appropriate decision node is selected and its text is compared with the new text. If not equal the high-low decision selects the appropriate path into the tree and selects the next appropriate file record for comparison. If no directional location is found, a new node is created and the record is added to the file at the end. The appropriate information concerning its location is inserted in the prior node and the control counter which provides continuous locations for the system is augmented.

The tree file has several advantages—one of which is the conservation of storage space. Records can be blocked as fixed or variable length, and packed to maximum hardware specifications. Retrieval speed is I-O bound and lapsed time is a function of the number of nodes to be stepped through to reach a final decision. Since the process is effectively binary, each decision node eliminates one half of the remaining file records if the tree is perfectly balanced. Each segment of the MARC record can be separated as a discrete file or the entire record can be treated as a single file. Exploded records require linkage to the remaining segments of other files.

The primary disadvantage of the tree is the maintenance problem. Periodically the entire file must be rebalanced and the starting nodes

reevaluated to provide the necessary balance. Reloading a structured tree file of a million bibliographic records represents a difficult and time-consuming task even for a computer.

THE MULTIPLE RING

The multiple-ring idea has some similarities to the structured tree, particularly to the alphabetic appearance of an unsequenced file. However, the location reference of the file record has different functions. Location A refers to the location of the record which alphabetically precedes the current one. Location B refers to the location of the record which follows the current one. Location C is a lateral reference to the next ring.

Multiple-ring filing has the same complexity as the structured tree. A binary search is made of the file until a position is found for the record. The location references of the preceding and trailing records are modified to fit the inclusion of the new record, and the locations of those records are added to the current record which is added to the end of the file.

The file maintenance problem is eased from that of the tree because the file is based upon its contents and not what the contents may eventually be. Periodically the ring files must be regenerated to maintain binary search speed. The reloading of such files is much less complex.

One of the primary advantages of the ring is the conservation of space. Another is the retrieval speed, although it is also I-O bound.

The three techniques of arranging data for access have distinct utility in a library situation. Our experience in testing these different systems suggests that a completely on-line library system will have several methods of data arrangement and several file structures. The problem may not be in how to organize bibliographic data in files but how to provide adequate interplay among files of dissimilar order.

The programming required for these three systems is not difficult, but the ring and tree methods are extremely involved. Some of the basic tenets of each system were written in Autocoder for a 1401, others in COBOL for a Burroughs 5500. Consideration of the three systems brought about a study of a specific system for the Fondren Library which one of our wags called the "tree-ring" study.

THE INDEXED REGISTER SYSTEM

The register system incorporates components of the three systems described and adds some additional variations. The registry system is different primarily because all files are components of the

registry. The register file record contains all of the location data for all bibliographic elements exploded into other files plus some special data concerning each entry. The remaining files are indexes and authority files containing references to the registry records incorporating the text.

As mentioned, each file is an index and authority file, and in addition is a holding file. The main entry file is a good example of the multipurpose basis for the file. There are three record types—index, content, and holding. The index record includes a reference to the location of the authoritative content record in addition to the actual text of the index. The content record contains the complete MARC text which becomes the authoritative form for that entry. The holding record contains a reference to the location of the formal text plus multiple locations of registry records which utilize the authoritative text.

It should be noted that the main entry file will actually contain added entries as well as main entries. Each reference to the registry or from the registry will contain a flagging character to indicate the use of the authoritative text in a record.

The component files of this system are perhaps unusual because the actual text occurs one time and is not repeated. If the main entry, authoritative text must be altered to add a death date or a given name, in most instances the connection can be made without modifying any of the other records in the file.

Preliminary evaluation of the major data files of the library shows no balance between the three file arrangements in the registry system.

Hashed files clearly dominate the group followed by the rings. Tree structures are used where they can be advantageous.

In retrieving data in this system, the primary problem is to reach the register as quickly as possible. Once the register record is identified all of the component elements of the text can be drawn together for output—as a MARC II record. This exploded data system does have its drawbacks; it is expensive. Nevertheless, it does provide direct access interplay between files threading data laterally through the register and vertically in all files.

The retrieval technique in terms of multiparametered, multilogical sets of data can be organized for greatest efficiency for the computer and for simplest procedures for the user. There does not appear to be any way of avoiding a special language for using library files. We are now exploring these problems and soon hope to have a working on-line cataloging system.

A user could type the command "seek" and then specify the logic text type and text to be sought, for example, "Seek not title collected or selected author hemingway." The user is requesting all materials

Table 2. File Arrangements in the Registry System

File Name	Hashed	Tree	Ring
Current Loans	x		
Reserve Inventory	x		
Borrower File	x		
Holds	x		
Main Entry			x
Title		x	
Edition	x		
Imprint	Included in Borrower File		
Collection	x		
Series Notes			x
Notes		x	
Subjects			x
Added Entries			x
Added Title		x	
Added Lines			x
Call Number	x		
Card Number	x		
Order File	x		
Invoice File	x		
Process Control			x
R.I.C.E. Accounts			x
Fund File	x		
Gift & Exchange			x
Binding File			x
Statistics		x	
Register	Sequential only		
Search Code		x	

by Hemingway which do not have the words "collected" or "selected" in the title.

Ordinarily the not logic should be tested first; however, in this instance the author entry will be more specific more rapidly and consequently the Hemingway string of data records will be used to access the register which in turn provides the location data for the title file. Each title extraction tests the text for the undesired words, and generates an output record for items fitting the requirements. Free form input with defining labels or words appears to be the best method of providing user-system interchange at Rice University.

The concluding commentary that follows is generalized. One does not compress five years of experimentation into a few pages without losing most of the detail. The use of MARC data at Rice University has been unique because of the particular circumstances which have existed and continue to exist. The absence of disastrous crises in technical processing permitted a level of experimentation with MARC data others could ill-afford. Some of the efforts in constructing concordances, providing selective dissemination of information from *Chemical Titles* and building research files for authors such as Thomas Mann provided an opportunity to compare techniques and problems caused by format differences.

The Fondren Library may be the only library to have used three different computing systems and four different languages on MARC data, as indicated below.

1. IBM 7040	MAP and COBOL
2. IBM 1401	Autocoder
3. Burroughs B-5500	ALGOL and COBOL

Our experience suggests that the existing format is easily handled by any of the current languages and probably by any hardware configuration on the market. Naturally, each system and each language present problems in programming for MARC applications. We have had our share, too, but have been able to resolve them in one way or another.

There are two points that I believe are of paramount importance to library automation. First, there does not appear to be a "best" way to do anything; everything seems to be relative. What works beautifully at one library may not do so well at another. People in library automation are under constant pressure to implement this and study that. The one thing most desperately needed and rarely obtained is adequate time to think about the situation and to use ingenuity and daring which produce the great innovations.

An aspect of the present situation which may prove more than a little disconcerting is micrographic storage. For five years we have encouraged, beseeched, and nourished the project which has grown into MARC. Now on the horizon looms a technological advance for which no one is ready—micrographic storage. The on-line storage cost for one billion bytes is not a small sum for today's libraries. The development of computer output on microfilm, microfilm and microfiche retrieval systems, some of which are computer driven, is a clear warning not to move too far toward the on-line storage concepts I have described.

The micrographic industry (into which the computer manufacturers are moving) is at the beginning of a period of successive important advances. These advances must be carefully watched during the development of any computer-based library system. MARC is here and is usable in many aspects of library processes. Use it. Be brave—but do not be foolhardy.

Robert S. McGee
Assistant Systems Development Librarian
and
Robert C. Miller
Head, Acquisitions Department
University of Chicago Library

MARC UTILIZATION IN THE UNIVERSITY OF CHICAGO LIBRARY BIBLIOGRAPHIC DATA PROCESSING SYSTEM*

The University of Chicago Library bibliographic data processing system has been under development since July 1966, and has been in full production operation for ordering and cataloging functions since November 1968. The basic design concept is for an integrated file system in which a single machine-held record is created and maintained for each bibliographic item. Item records are variously processed for such products as acquisitions purchase orders, fiscal reports, catalog card sets, charge cards, and pocket labels. The system is built around an on-line master in-process file to which records and data can be variously entered as needed for ordering, order record update, cataloging, etc. At any time the master file contains records in many stages of completion; its current size is over 40,000 records. Altogether more than 100,000 complete machine-readable bibliographic records have been created through the system. These are stored in a historical file on magnetic tape, to which master file records are transferred when their processing is completed.

Over 100 operational computer programs written in basic assembly language have been implemented to date. The library system currently runs on the University Computation Center's IBM System/360 model 65 computer, which operates under System/360 operating system with MVT (multiprogramming with a variable number of tasks). On-line files are maintained on an IBM 2314 Direct Access Storage Facility. System products are printed on an IBM 1403 high-speed printer with a special library print train.

The utilization of MARC data in production operations is based

*The work reported here has been supported in part by a grant from the National Science Foundation, and is the result of combined efforts by the University of Chicago Library Systems Development Office, associated programmers, and library operational staff.

upon merging wanted MARC bibliographic data with corresponding local data (such as fiscal and vendor information) within master file records. The Chicago system, which was developed during the same period as MARC I, has its own record format design, tagging code structure, and data element definitions, although in many respects these are similar to MARC II. This requires that MARC data be converted to the Chicago format using a special translating program. Thus the basic steps for utilizing MARC records in Chicago production operations are: 1) convert MARC records and data element formats to those of Chicago, and 2) merge the converted MARC bibliographic data with Chicago processing data in master file records. Once MARC data have been transferred to master file records, these records together with those created wholly from local inputs can be processed by existing systems and programs for various outputs. MARC data are used in order generation, catalog card production, or both—depending upon the timeliness of individual MARC records.

Figure 1 outlines the Chicago system and the use of MARC data within it. Virtually all acquisitions purchase orders, book charge cards and pocket labels, and Roman alphabet cataloging are handled by the system. Source documents containing data to be input for these products enter the Library Data Processing Unit (LDPU) from various technical processing departments. The LDPU, which consists of six full-time clerical equivalents, inputs these data off-line into paper tape on three IBM 1050 terminals. These tapes are read on-line from two of the terminals to the master in-process file located at the University Computation Center. Library products are printed in batches at night and are delivered the next morning to LDPU where they are inspected before they are routed to appropriate departments.

Utilization of MARC in this system is presently based upon matching available MARC records to ordering and cataloging source documents as they flow into LDPU. As indicated by Figure 1, records from new MARC tapes are added to the CUMARC file, which currently is a cumulation of MARC records from the last thirteen tapes, converted to the Chicago format. The LC card numbers of these incoming records are also added to the CUMARC index, which is a disk-held index by LC card number to MARC data available in the CUMARC file. The only automatic access to CUMARC records is through LC card number searches of the CUMARC index. The LDPU inputs LC card numbers from production source documents for nightly searching and receives a printout report of results on the following morning. Source documents for unmatched requests are put into normal production streams for local keyboarding of all needed data. Source documents for matched requests are separated for special inputs later. Matched requests cause updates of the CUMARC index that trigger automatic transfers of wanted MARC data from the

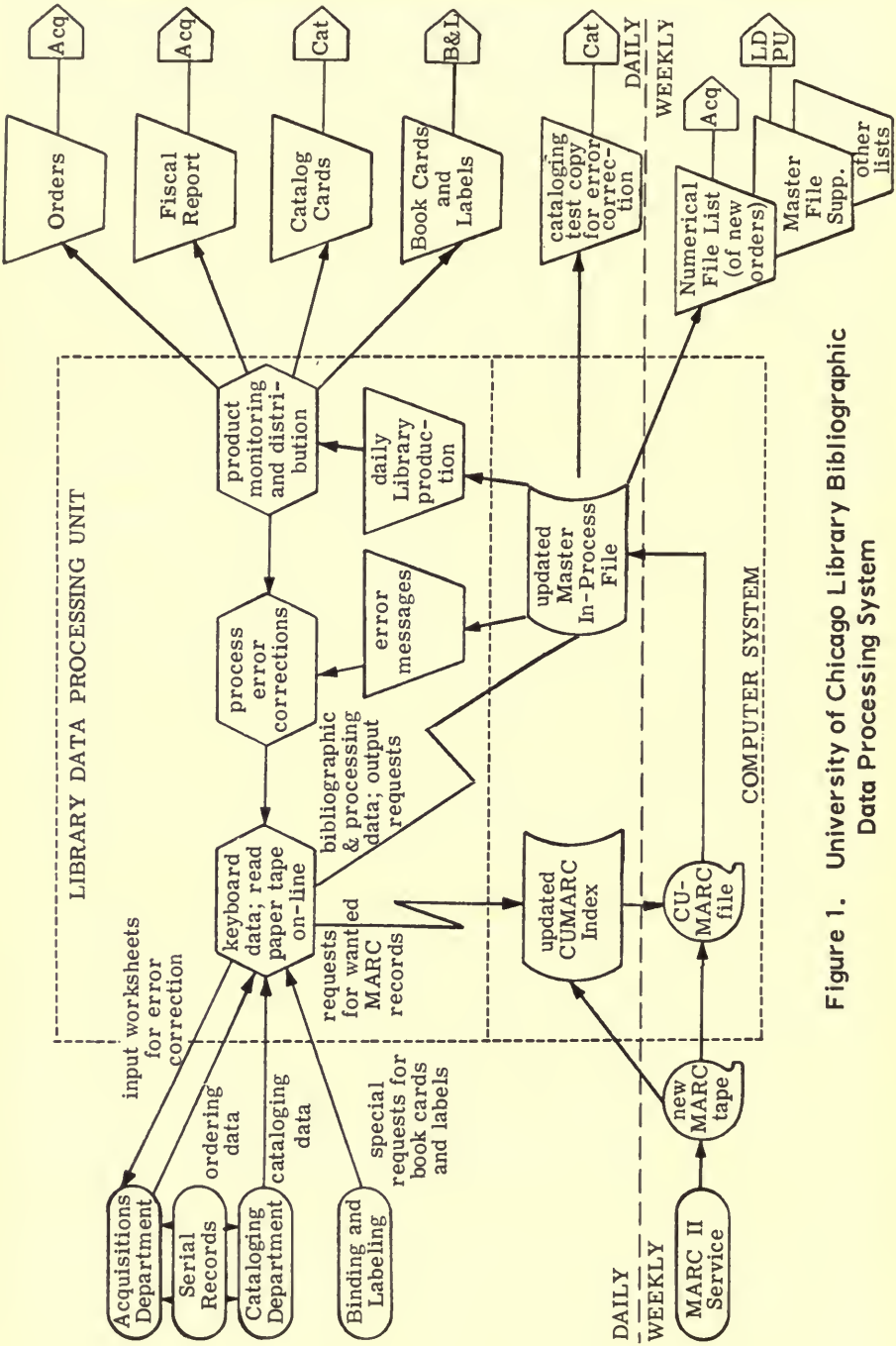


Figure 1. University of Chicago Library Bibliographic Data Processing System

Table 1. Outline of MARC Processing

Weekly New Tape Processing:

1. a) Convert variable length (ASCII) MARC II format records to 200-character (EBCDIC) University of Chicago format records. These records are called CUMARC records. At the same time drop all juvenile records, and within accepted records, all nonrelevant tags and data.
   b) Sort the output file into ascending order by LC card number.
2. a) Add new records to the cumulative MARC II file (CUMARC file) and to the cumulative index (CUMARC index).
   b) Move to a CUMARC historical file any records received prior to the last thirteen MARC tapes. Delete those records marked FF in the index.

Receive and Search Requests for Wanted MARC Records (Daily):

1. Check incoming data stream for item records containing tag 023 and LC card number of wanted MARC record.
2. Create an entry in the output request table from the item record number and the action code (in this case tag 023).
3. a) Read the output request table for requests for MARC records, get LC card numbers from master file records, and create an entry for each request.
   b) Sort the created entries into LC card number order.
   c) Compare entries with CUMARC index; merge matched entries with index; print out unmatched entries as the non-match report.

Move Requested Records to Master In-Process File (Weekly):

1. a) Read index to find records needing action. Read MARC record, read MASTER record, and merge the data. Write output on tape. Merged records are marked FF in CUMARC index.
   b) Print tape (optional).
2. Add new merged records to master in-process file.

For Production Any Time after a CUMARC/Master Record Merge:

Update master file record with data which are required for production, and which act as production signals by creating entries in the output request table.

CUMARC file to master file records during the next MARC production run. Precedence rules determine which MARC data elements overwrite any corresponding data that may have already been locally input. MARC bibliographic data generally take precedence, although some local data elements such as list price and call number have priority. After MARC data have been merged with master file records, these are updated by final inputs of local data needed to complete the records before production. Inputs of these data also serve as signals for wanted processing. For example, in cataloging, the input of a card distribution code specifies the cards to be printed from a record, and acts as a trigger for catalog card production.

The source documents currently used in this matching procedure are LC cards and proofslips with the MARC acronym which indicates inclusion in MARC II. The library receives a set of proofslips for use in selection review and ordering, and a set of LC cards used primarily to supply Library of Congress data for cataloging. Consequently these appear as ordering and cataloging source documents in the Library Data Processing Unit production streams, and are processed daily through the automatic search procedure.

Identifying the LC card numbers of wanted MARC records from LC copy has permitted the use of MARC data where possible, and has allowed alternative manual inputs from these source documents when matches have not been made. This has provided two basic advantages: the avoidance of cumulating backlogs of unmatched items awaiting MARC data; and the opportunity to automatically search the CUMARC index for wanted LC card numbers, which has also enabled postponing special printouts for access to MARC records until operational needs could be better judged.

Matching procedures have also been conducted for English-language books received through the library's acquisitions approval program with a book jobber, and for English-language cataloging storeroom items awaiting Library of Congress data. Where available, LC card numbers of these items are input for automatic searching. For matched requests, MARC data are transferred to master file records, and main entry cards are printed (a standard system capability). These cards are matched to corresponding items, which can then go immediately to cataloging. When the cataloging source documents reach the Library Data Processing Unit, only updating inputs of necessary local data, as above, are required for the production of catalog card sets.

From the outset we have viewed these MARC activities as primitive and initial—not final—production operations. Our approach has been to get quickly into MARC utilization using available system capabilities and avoiding large outlays of special programming where possible, in order to gain basic operational experience before further developments. Within the scope of our current design and operations

two factors significantly limit MARC utilization: 1) the relatively small machine-held CUMARC index, limited to entries for only those records received on the latest thirteen tapes; and 2) an automatic query capability which requires the LC card number of a wanted MARC record to be input each time it is searched (there is no capability for automatically repeating searches for unmatched requests). Although the inherent limitations of this scheme were realized, we judged it better to begin early operations with their projected limits rather than delay initial MARC utilization until the development of more sophisticated capabilities. As a result, our MARC activities have provided some solid operational benefits, and given us the experience we felt necessary for further development.

DATA ON THE FIRST THIRTY-TWO WEEKS OF MARC UTILIZATION

Tables 2 through 5 present data on Chicago's first thirty-two weeks of MARC production operations. By the end of this period (which coincides with the receipt of tape 52 of MARC Volume 1*) 5,959 MARC records had been used in ordering and cataloging, as shown by Table 2. Since 1,536 records used for ordering were used again at the cataloging stage, total uses of MARC data in these two functions came to 7,495. Further repeated uses in cataloging are anticipated for another 500 records used for orders still outstanding, bringing expected uses of records from the first fifty-two tapes to a minimum of 8,000. It is likely that still more new uses of these records will occur until tape 52 is phased out of the system.

As might be expected, initial usage was affected by start-up problems which included: a four-month lag between the receipt of tape 1 and the beginning of operations, by which time nineteen tapes had been received; the relatively short life of some records in the system as the backlog of tapes was processed through rather quickly (e.g., records from tapes 1 through 9 were in the system less than five production weeks); and problems encountered in reading tapes, some of which were added later than desirable, and four of which were not added at all. Table 3 gives cumulative distributions of MARC usage during the first thirty-two weeks of production, which reflect increasing uses as initial troubles were overcome.

Table 4 presents data on CUMARC index searching operations. Searches conducted with LC card numbers taken from LC copy with the MARC acronym were significantly more successful (58.6 percent matched) than searches made with LC card numbers taken from

---

*Volume 1 of MARC consists of fifty-four weekly tapes; fifty-two weekly tapes were received during the first year of MARC II distribution.

Table 2. Utilization of MARC II Data during the First Year of Distribution

The University of Chicago Library began production utilization of MARC II data in August 1969. Tape number 52 was received on March 24, 1970, completing the first year of MARC II distribution. MARC usage during the period from August 6, 1969 to March 27, 1970 is shown below. Since the CUMARC file is a cumulation of records from the latest thirteen tapes, some records from the first year's tapes will remain available for use in the system until tape 52 is phased out in another thirteen weeks.

	Number	Percent	Number	Percent
Records used for ordering			2,061	(34.6)
Records used for cataloging				
Matched to items received through acquisitions approval arrangement	622			
Matched to cataloging storeroom items awaiting LC data	432			
Matched to cataloging source documents (LC copy with MARC acronym)	2,844			
Total used only for cataloging	3,898	(71.7)	3,898	(65.4)
Records used for cataloging that were previously used for ordering	1,536	(28.3)		
Total	5,434	(100.0)		
Total records used in production			5,959	(100.0)

Total uses of records

	Number	Percent
for ordering	2,061	(27.5)
for cataloging	5,434	(72.5)
	7,495	(100.0)

Table 2. (continued) Utilization of MARC II Data during the First Year of Distribution

Records from tapes 1-52	
received	50,234
not processed (four defective tapes)	-3,320
processed	46,914
automatically deleted (estimated at 6%)	-2,814
added to CUMARC file (estimated)	44,100
used as of March 27, 1970	5,959
(11.9% of 50,234 received)	
(12.1% of estimated 46,914 processed)	
(13.5% of estimated 44,100 added)	

books (21.2 percent). Book searches were made only for items falling within MARC's defined limit to English-language materials. Since the MARC acronym is not given with LC card numbers appearing in books, there has been no other convenient way to selectively identify and search numbers with higher matching probabilities. It is suspected that many of the books that were searched had been cataloged by the Library of Congress prior to MARC II or before it became comprehensive, and therefore were not covered by it.

Searches based upon card numbers taken from LC copy depend essentially upon the extent to which these appear as source documents in the Library Data Processing Unit ordering and cataloging production streams. Virtually all ordering is handled through the bibliographic data processing system and, as shown by data from a recent six-month period, 87 percent of the cataloging production. The 13 percent handled otherwise is largely cataloging in non-Roman alphabets which falls outside the character set of the library print train and cannot be printed by it. Sixty-five percent of the total cataloging was from Library of Congress data. No information was collected on the distribution of LC data among Roman and non-Roman alphabet cataloging records.

Table 5 shows the composition of source documents received by the Library Data Processing Unit, and roughly indicates the potential that existed for using MARC in the described system during these periods. For ordering, 9.1 percent of the source documents were LC copy with the MARC acronym. Assuming that corresponding MARC records were received, a perfect system could potentially have matched MARC data to this percentage of ordering. It is

Table 3. Cumulative Distributions of MARC Usage during the First Thirty-Two Weeks of Production Utilization (August 6, 1969 to March 27, 1970)

	ORDERING		CATALOGING			
			Total		Previously Used for Ordering	
Production Week	Records	Percent	Records	Percent	Records	Percent
1			38	1		
2			44	1		
3			75	1		
4			75	1		
5	86	4	168	3		
6	100	5	260	5		
7	104	5	473	9		
8	107	5	577	11		
9	283	14	709	13		
10	352	17	793	15		
11	453	22	996	18	60	4
12	474	23	1191	22	101	7
13	656	32	1421	26	146	10
14	682	33	1600	30	209	14
15	761	37	1783	33	275	18
16	808	39	1937	36	334	22
17	840	41	2089	38	385	25
18	868	42	2303	42	542	35
19	954	46	2606	48	621	40
20	1071	52	2793	51	687	45
21	1136	55	3022	56	779	51
22	1200	58	3224	59	826	54
23	1369	66	3530	65	953	62
24	1454	71	3802	70	1027	67
25	1534	74	4073	75	1100	72
26	1591	77	4226	78	1162	76
27	1846	90	4445	82	1238	81
28	1934	94	4612	85	1275	83
29	1950	95	4891	90	1387	90
30	1969	96	5056	93	1445	94
31	1993	97	5239	96	1476	96
32	2061	100	5434	100	1536	100

Table 4. CUMARC Index Searching Results

The CUMARC index is an LC card number index to records from the latest thirteen MARC tapes that are available on the CUMARC file. Automatic searches of the index are performed upon inputs of LC card numbers for wanted records. The current design requires that a given request be input each time it is searched; there is no capability for automatically repeating searches for unmatched requests, and unmatched LC card numbers are not re-input for repeated searches. Requests are only matched if wanted records are available at the time of request.

Searches	Requested	Matched	Percent of Requested
Ordering (9/11/69 - 3/27/70)	3,252*	2,061	63.4
Cataloging			
For acquisitions approval items (8/6/69 - 1/29/70)	2,259†	622	27.5
For storeroom items (8/6 - 1/29)	2,705†	432	16.0
For regular cataloging source documents (9/11 - 3/27)	5,112*	2,844	55.6
Total	13,328	5,959	44.7
*Total searches of LC card numbers taken from LC copy with the MARC acronym	8,364	4,905	58.6
†Searches of LC card numbers taken from English-language books	4,964	1,054	21.2

believed, however, that additional opportunities for MARC usage may have existed, since the composition of source documents for priority ordering was not determined. It is suspected that as much as 10 percent of these may have been LC copy.

The MARC potential in cataloging is represented by MARC data carried from ordering (all of which were used) as well as by LC copy with the MARC acronym. Together these constituted 26.6 percent of the cataloging total, which interestingly is nearly half the LC copy received for cataloging during this period.

By comparing data from Tables 3 and 5, the performance of the

Table 5. Composition of Ordering and Cataloging Source Documents Received by the Library Data Processing Unit

The library receives a set of LC proofslips for use in selection review and ordering, and a set of LC cards to supply data for cataloging. This tabulation shows the extent to which these LC copies appeared as ordering and cataloging source documents in the Library Data Processing Unit, and of these, the proportion with the MARC acronym.

Ordering September 11, 1969—March 27, 1970		Number	Percent of Total
From LC copy without MARC acronym		3,553	(10.0)
From LC copy with MARC acronym		3,252	(9.1)
From other source documents			
manuscript cards	15,398		
bibliographies	4,687		
for priority ordering	8,692		
	28,777	28,777	(80.9)
	Total	35,582	(100.0)

Cataloging October 13, 1969—March 27, 1970		
From MARC data previously used for ordering	1,536	(7.2)
From LC copy with MARC acronym	4,171	(19.4)
From LC copy without MARC acronym	5,948	(27.7)
From other source documents	9,814	(45.7)
Total	21,469	(100.0)

---

described system in matching MARC to production needs can be evaluated. Table 5 indicates a MARC potential of at least 9.1 percent for ordering from the source documents with the MARC acronym. Table 3 shows that 2,061 MARC records, equivalent to 5.7 percent of the ordering source documents, were actually used. Similarly, for a given period in cataloging, 26.6 percent of the total source records were either MARC data previously used for ordering, or LC copy

with the MARC acronym. Data from Table 5 for this same period show that 4,171 MARC records, or 19.4 percent of the total, were actually used for cataloging. (It should be noted that all 7.2 percent of the data carried from ordering were used, making the use percentage favorably higher).

Our conclusion is that we can make more effective use of MARC data, especially in ordering. Considerations of ways to improve matching MARC data to production needs are discussed below.

PLANNED CHANGES IN CHICAGO'S SYSTEM FOR MARC UTILIZATION

The fundamental problem of MARC utilization is the matching of MARC data to production needs. A given MARC record may arrive before or after the need for Library of Congress data is established, or a corresponding MARC record may never arrive at all. A system covering these possibilities could be one with capabilities to 1) compare incoming MARC records to known production needs, and match them where possible; 2) to store unmatched records and provide accesses to them that will permit subsequent matches; and 3) to permit requests and matches of MARC data at any time, and after an appropriate waiting period to identify unmatched requests and to take remedial action. This essentially is Chicago's approach to expanding its capabilities for MARC utilization beyond those already described.

MARC's role in production operations may also be characterized as active or passive. A passive use is one in which the production need is established before the availability of MARC data is known, and to which they are subsequently matched. An active role is one in which the presence of a given MARC record is known before the need for it is, and which itself stimulates its own use, as in the case of MARC record printouts from which ordering selections are made. Both modes of utilization are incorporated in Chicago's planned expansions.

These plans provide for expanded CUMARC index and querying operations in which unmatched requests are retained and automatically searched against incoming tapes, so that wanted records are immediately identified and matched to production needs. Unmatched records from the incoming tapes will be processed through various deselection routines to eliminate categories of records with relative low usage probabilities. The remaining records will be examined by various data elements for printout and routing to special processes. Most of these will be divided into groups on the basis of their LC subject classifications, and sent to appropriate selectors for acquisitions review. Records that are subsequently selected for ordering through this process can then be entered into the normal production stream where they will be identified as available MARC

data. Record printouts not selected at this time will be filed by title into the manual ordering file. Here they will be accessible by title searches of order requests (a standard procedure) to which MARC data can thus be matched. Since this file is updated upon the receipt of items, these printouts of MARC records can also be immediately matched to received material that will go next to cataloging. This method of accessing MARC data by title will be used until such time as superior means are available.

It is intended that printouts of MARC records will replace their corresponding LC copy in library processing. LC cards and proof-slips with the MARC acronym will no longer be used, although the non-MARC copy will continue to be used. The consolidation of all uses of Library of Congress data for MARC scope materials into a single set of MARC printouts should reduce problems of duplicate order requests and other unwanted multiple uses of LC copy that have been experienced in the past.

At an estimated arrival rate of 1,200 records per week, MARC utilization leads quickly to problems of file size and storage. It becomes important to seek usage of individual records as quickly as possible, to reject incoming records that are unlikely to be used (if these can be identified), and to purge unused records from the system after a suitable time. To maintain the size of the CUMARC file within reasonable limits that will still allow unused records to remain on it for perhaps as long as six months, we have developed deselection algorithms which have just recently entered testing for implementation with the expanded system. In CUMARC operations to date the only records automatically dropped from incoming tapes have been those with the juvenile indicator in the intellectual level code, and correction records that update previously received data. These deletions have amounted to 6 percent of the total records received. We are now evaluating rules to deselect records for items with imprint dates two years old or earlier, and also to eliminate records falling within certain Library of Congress classifications. For some classifications containing large numbers of records of marginal utility, we are considering print and purge algorithms that will provide printouts (for manual file storage, to preserve data availability) and then drop machine records from further processing to save file space.

After incoming MARC records are compared to the CUMARC index for identification of needed data, and deselections have been made for unwanted records, printouts of records for PL 480 items (as known from their LC card numbers) and for items from selected national bibliographies (identifiable from codes in MARC data element 015) will be made for special handling. The PL 480 printouts will go directly to the South Asia, Arabic, and Slavic storage areas so that cataloging data can be supplied for items which are normally

received in advance of corresponding LC records. Printouts for national bibliography items will be compared directly to the ordering file, where it is expected that a significant number will match orders that have already been made from these bibliographic sources, which generally precede the arrival of corresponding Library of Congress data.

The remaining records from the incoming tape will then be sorted according to a detailed machine-held table of LC classification schedules, for printouts to be distributed among the library's selection staff. The development of specifications for this table has proved to be one of the most difficult and time-consuming aspects of our work on MARC utilization. The actual programming was not particularly difficult. The problem has been to decide which selector should be assigned which classifications. The selection staff is based variously on subject disciplines, area programs, and administrative units. Overlaps among selection responsibilities are frequent, both in areas of collection building and personal interest. It was desired that division of the classification schedules should be complete, but this was difficult to achieve. For example, proposed assignments of the *N* schedules (for Art) at one point involved four pages in which almost every successive number for some areas went to a different selector among the five responsible.

Most of the suggestions received from selectors were not unreasonable in themselves. In combination, however, they were formidable. But after much discussion and debate, sufficient compromises were reached, and specifications for special MARC selection printouts were given to the programmers.

These printouts may significantly improve utilization of MARC data at the ordering stage. Looking at Table 5, the category of ordering source documents called manuscript cards constitutes 43 percent of the total. Since manuscript cards are standardized forms generally used to transcribe ordering data from non-LC sources such as vendors' catalogs, blurbs, bibliographies, etc., a large number of the items they represent may fall within MARC scope. It is hoped that thorough utilization of MARC printouts will replace some of these selection sources, and it is believed that MARC's timeliness enhances this possibility. Comparisons by the Library Systems Development Office of LC proofslip and MARC record arrival dates during a fifteen-week period from April through July 1969, show that 80 percent of proofslips with the MARC acronym matched MARC records received the same week or earlier.[1] The conclusion of this study is that MARC records are timely enough to be used in the selection and ordering processes where proofslips are now used. It is also hoped that the effectiveness of ordering from MARC printouts will be great enough to reduce ordering from other sources, even though some of these may be received sooner than corresponding MARC records.

Although detailed specifications for upgrading CUMARC index operations have not been completed, basic objectives have been outlined. Unmatched LC card number requests for wanted MARC records will be retained and automatically searched against incoming tapes to identify needed data as soon as possible. Requests will be entered with simple codes indicating desired use, so that matches can result in appropriate printouts and/or automatic transfers of MARC data to master file records. Printouts will be in labeled, ordered arrays to facilitate matching them to specific library locations and needs. It is planned that after suitable waiting periods, unmatched requests will be reported so that alternative manual processing may be taken as needed. Since the CUMARC index will retain status codes for CUMARC file records, a further opportunity exists to report multiple requests for given records, and check them if necessary for unwanted ordering and cataloging duplications.

We believe these index capabilities will be particularly effective in matching MARC data to cataloging storeroom books. Many of these are received as standing, continuing, and blanket-order items for which separate orders were never placed. The opportunity to request cataloging data by their LC card numbers should enable matches to MARC records for which printouts are still in selection processing, as well as to records yet to be received.

The changes outlined here are not regarded as final solutions to MARC utilization. As MARC grows, we expect that our system for using it will also.

SOME GENERAL COMMENTS ON MARC UTILIZATION BY AN ACQUISITIONS LIBRARIAN

From the standpoint of an individual library, MARC II Distribution Service is the dissemination on magnetic tape of standardized machine cataloging data for materials being currently cataloged by the Library of Congress. For the present, at least, its scope includes only English-language monographic titles. With this in mind, a number of specific factors related to MARC can be usefully considered, although definitive statements in this area may not be possible.

Reliability is an important element in any ongoing system, especially one that is machine based. Is MARC reliable enough for a library to base its processing operations on it? There are, in addition to general problems of data content reliability, at least three other aspects of reliability involved. Politically, is MARC secure enough in the range of services offered by the Library of Congress to be relied upon, or is it susceptible to cutbacks or discontinuance? While this seems unlikely, it is far from impossible in an era of pressing federal budgetary problems. The possibility presents a double-edged dilemma for individual libraries. If libraries do not

make substantial use of MARC, it may be viewed by budgetmakers as expendable. On the other hand, libraries that build an operational reliance on MARC could be in serious difficulty if the service were curtailed. One must also ask if MARC is physically reliable enough. The tape problems of the first few months are apparently a thing of the past, but this kind of problem, especially of a less obvious nature, could happen again, and could make serious trouble for a library system built around MARC, particularly if garbage records were fully entered into the local system prior to detection.

Thirdly, is the MARC format itself stable enough or are there likely to be changes that could have serious repercussions on a user library's systems? The new fourth edition of the MARC users guide dated April 1970[2] cites a large number of changes since the third edition of March 1969. Definitions and structures of individual data elements, the addition or deletion of elements, modifications in the structure of the leader and directory may at first seem to be relatively insignificant changes. Each change must be carefully examined, however, for some may require extensive software maintenance and data base updating to maintain both efficient operation and consistent data base quality. For libraries operating with severely restricted budgets for systems and programming work, even few such changes could prove critical. Realistically, one cannot expect a radical new service such as MARC II to remain static. Any library that plans to use MARC must make allowances in staffing to handle the inevitable changes and systems maintenance.

Another aspect of the MARC service significant to library operations is the timeliness of record arrival dates, whether they are used in selection, ordering, or cataloging. It is important to bear in mind that, glamour notwithstanding, bibliographic data a year and a half old are just as useful or useless on MARC tapes as the same data in printed card form. Speed is especially important for book selection purposes, and unless the speed of the MARC service can be made faster than has been true for printed cards, much of the data will not be timely enough for the selection and ordering functions in a research library. One theoretical processing goal, which will probably never be attained, is to select, order, and catalog from MARC data all materials in MARC scope obtained for the library. The more quickly the data come, the closer this goal can be approached. Indeed, speaking only as an acquisitions librarian, speeding up English-language data would seem preferable to expanding the scope of MARC. Since separate administrative units of the Library of Congress are involved in this particular problem, the alternatives may not be real trade-offs.

In any event, for selection purposes at least, non-MARC sources, even for MARC scope material, will undoubtedly remain in use indefinitely. MARC is, after all, first and foremost cataloging data,

and its prime usefulness locally will probably be for cataloging purposes, as is generally true for the printed format of these data. Nonetheless, given sufficient volume and timeliness, MARC data can be efficiently utilized at the selection and ordering stages.

Consideration of local MARC implementation raises questions on the quantity, both relative and absolute, of MARC data that can and will be utilized by the individual library. Specifically, questions about the proportion of the library's acquisitions and cataloging within MARC scope and the number of MARC records the library will potentially use, both absolutely and within given time periods, become critical. The answers to such questions depend in part on local library policy. To realize the full potential of MARC, a library must make conscious efforts to utilize the maximum possible number of individual records as much as possible in both ordering and cataloging. In some cases this may require processing delays, but it need not imply that the material would be less quickly available to the library's users. For routine materials not in special demand, the superior quality of MARC data and the relative ease of processing with this data may justify waiting for carefully determined periods of time for the arrival of MARC records before ordering. Where user demands make this delay unacceptable, minimum non-MARC order data could be input manually to get out orders, and the material circulated on arrival without formal cataloging. Use of the material "uncataloged" could thus continue, with full cataloging delayed until the appropriate MARC record arrives.

The quality of the data received and used must also be taken into account. Can an individual institution accept without modification the Library of Congress's choice of entries, subject headings, classification, descriptive cataloging standards, etc.? Traditionally librarians have had an apparent compulsion to modify most Library of Congress records in some way, however minor. In a manual operation, the costs of such modification tend to get buried. In a machine system, while modifications are not necessarily difficult, their costs tend to be more obvious. A basic economy implicit in an externally generated machine-readable data base such as MARC is that the data will be modified as little as possible. The real question then becomes whether the modifications are of enough significance to justify the costs. To modify MARC data without good reason is to endanger the cost advantages of local utilization. If a given library has in the past been unwilling or unable to accept Library of Congress cataloging, MARC may be difficult and unreasonable to adopt unless the library is prepared to re-examine its basic policies and, where necessary, to redesign its systems to take efficient advantage of MARC.

This raises seldom-answered questions of MARC costs. Little hard data exist in this area, apart from the known $800 a year subscription fee. To this must be added costs of systems development

and the various continuing costs of operation and maintenance. A MARC processing system will probably be costly to set up and operate. This is not to say there are not potential long range economies, only that they may not be immediately evident. Moreover, the library's policies will drastically affect whatever economies will be obtained. Where can the economies come from if they do come? There are basically at least three specific areas of potential saving. First, there can be significant savings in input costs. Since only local order data elements such as "fund" and "vendor" need be manually input, using MARC data for order generation significantly reduces the required manual input. Given the much larger data requirements for catalog records in contrast to orders, the reductions in input for catalog card production should be even greater. To realize these savings on a net basis, it is necessary, of course, to develop the over-all operation so that time saved in input is not used up in other preliminary tasks. Secondly, the use of MARC data reduces and in some cases may eliminate the need for proofreading, a costly and frustrating task in any system. Thirdly, the use of MARC input at the order stage cuts down significantly the amount of updating that must be done when the material is actually being cataloged. In general, an important benefit of a machine-readable data base is the repeated use of the individual record. Both the long-term and individual use costs should decline with each use. It thus behooves libraries to make maximum use of MARC data for all aspects of selection, ordering and cataloging.

One last facet of extensive MARC utilization—the developmental cost of a computer-based system and the relative economies that should result from the use of MARC versus non-MARC data in carefully designed operations, together with the increasing cost consciousness of library administrators—suggests that the MARC usage will affect library policies and procedures to an extent never realized in manual systems. This may inevitably make libraries individually and collectively more dependent than ever on the Library of Congress. While this may be altogether fitting, it does suggest certain caveats: first, the library community must develop new methods of participating in the determination of Library of Congress policies relative to MARC, and secondly, in designing its processing systems the individual library must maintain a large degree of flexibility. For example, while seeking and preferring MARC sources for ordering, large research libraries at least must avoid the probably inevitable tendency to slight or refuse handling non-Library of Congress sources for selection.

The library contemplating the use of MARC can view the service in at least two different ways. On the one hand, stress can be put on the fact that these are the same data previously available in LC cards and proofsheets. In itself this is true. MARC data are Library of

Congress data, and as such are subject to many of the problems traditional with Library of Congress cataloging, plus some new ones. Seen in this light exclusively, MARC utilization may not be particularly attractive. Considered from another viewpoint, however, the format is what becomes important. Here the format is not a mere accident; the medium is the message. The concept of distributing data in this format is so radical that it requires rethinking of each individual library operation as part of a total system.

It is with this approach that the planning of MARC usage is most exciting. The challenge is not one of automating existing procedures, but of making the maximum utilization of the various data elements in the MARC record and the machine manipulation which these elements permit, thus developing new services and new approaches to library operations. For example, automatic generation of a full set of library records upon the arrival of MARC data may be feasible. If a library has a standing order for all publications of a given publisher or in a specified series, and the library will use Library of Congress cataloging and classification completely as is, a full set of catalog cards, book cards, labels and other related products would be produced immediately upon receipt of the MARC record regardless of the receipt of the material. When the MARC record arrives before the actual material, automatic claims to vendors might be generated. It may even be possible to automate certain limited areas of book selection. Orders could be generated on the basis of publisher, series or various combinations of selected data elements such as level, subject headings, LC classification, etc. The presence of bibliographic (list) price in many records opens the door to automatically utilizing these data for bookkeeping operations. It will also permit the automatic generation of studies on book publishing and pricing by subjects, publisher or other elements, a desirable feature for collection development and budgetary planning.

Even without automating the actual decisions, however, MARC offers great potential for the selection process. Selection tools could be custom-made for individuals by subject or area, eliminating or specially marking those items for which series, standing or individual orders are outstanding, or which are routinely selected in other media, as with PL 480 or national bibliography sources. Multiple copies of selection lists could be provided to make selection easier and more broadly based. On such lists individual records could be simultaneously referred to a number of selectors. This is a desirable feature not readily possible with the routing of single copies of proof slips, a common procedure in many large libraries.

After selection MARC has the potential of revolutionizing order techniques. The presence of the Standard Book Number (SBN) as a MARC data element and its growing acceptance by publishers and jobbers offer the possibility of ordering from vendor via teletypes as an alternative to the mailing of printed orders.

While time does not permit considering them in detail here, there are certain additional long range implications of MARC which merit careful examination by library planners and administrators. First, with external machine-readable data available, the development of local computer-assisted general library processing systems becomes more attractive economically. Thus at least one commercial firm is talking of basing its services with MARC as its primary data base.[3] Secondly, the utilization of machine records and their attendant systems will affect the type of work in library processing operations and ultimately require long range changes in the administrative organization of the library. Already the limited operations at Chicago are making clear that new skills and revised work flows are required for efficient and economical operation. Finally, the widespread acceptance of the MARC format has been one directional but as a communication medium, MARC can ultimately be utilized for the transfer of information to Library of Congress as well as from it.

In sum, the MARC II Distribution Service offers the potential for the radical transformation of many library operations. The limits to this will be set in good part by the resources available for planning and by the imagination and skill of library personnel.

REFERENCES

1. Payne, Charles T., and McGee, Robert S. "Comparisons of LC Proofslip and MARC Tape Arrival Dates at the University of Chicago Library," *Journal of Library Automation*, 3:115-21, June 1970.

2. U.S. Library of Congress. Information Systems Office. *Books: A MARC Format*. 4th ed. Washington, D.C., Library of Congress, 1970.

3. Pearson, Karl M. *MARC and the Library Service Center: Automation at Bargain Rates*. Santa Monica, Calif., System Development Corp., 1969.

Fred C. Cole, Jr.
Programmer
Autocomp, Inc.
Bethesda, Maryland

COMMERCIAL USERS OF MARC: AUTOCOMP, INC.

In 1966 Autocomp was founded as an information systems and computer software company to work principally in the text data input, manipulation, full-text retrieval, and photocomposition areas. Autocomp developed an operational system called RECOMP (Retrieval and Composition) which at the time of this writing had processed over 325 million characters of data. Prime emphasis has been the computerization and retrieval of legal information and the manipulation of data for the codification of laws. However, in the latter part of 1968 it was recognized that RECOMP could operate equally well for the input, manipulation, and output of bibliographic data; work with the MARC tapes of the Library of Congress was begun. Since RECOMP is a broad, universal system, easily adapted to all forms of text processing and because it is fully operational, we were able to process the MARC data and to make it easily accessible.

A general description of RECOMP and its utility for the processing of bibliographic data follow. The RECOMP system is divided, conceptually, into five major areas: 1) input, 2) maintenance, 3) manipulation, 4) composition, and 5) retrieval (see Figure 1). Each major area comprises several program modules. The following remarks are devoted to a description of these programs by area, with particular attention given to concept as opposed to technical detail.

INPUT

MARC data can be supplemented through local input, utilizing the input portion of the RECOMP system. The RECOMP method allows us to accept data from many different input devices of which DURA and Friden paper tape, MT/ST, and the REI optical scanner are a few. Fixed as well as variable data can be processed. Data elements must be explicitly identified whenever meaning is not implied. For example, the system is able to accept tapes constructed on other computer systems and to identify data elements through whatever means were employed by the other system, such as fixed record positions or imbedded mnemonic identifiers. If data is to be keyed

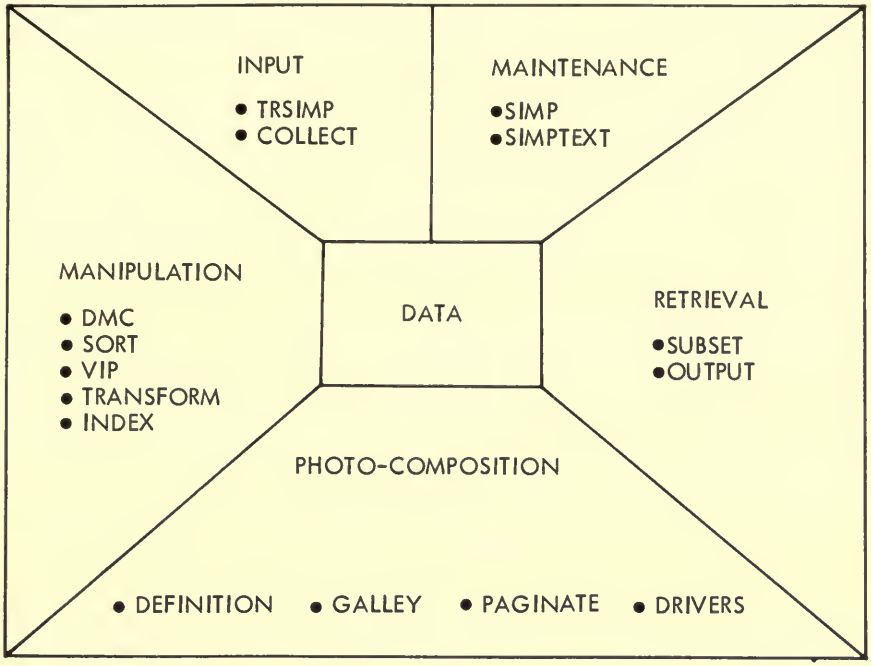


Figure 1. The RECOMP System

explicitly for manipulation, we would need to imbed mnemonics if the format is insufficient to identify the elements of input. All input is considered to be raw and to require restructuring. To accomplish this our transaction and structuring program is employed.

GENERALIZED DATA TRANSLATOR (TRSIMP)

TRSIMP will translate any bit structure (1 or 2 byte) into IBM EBCDIC. Structuring capabilities of this program enable the system to output a file that can be updated on either a line-by-line or paragraph-by-paragraph basis. Line-by-line updating is preferable if the input has little or no identification. Paragraph-by-paragraph updating is preferable if Autocomp has input the data itself. If the input data must go through an extensive structuring process, TRSIMP is augmented by a data manipulation command deck, which I will discuss later.

MAINTENANCE

The maintenance portion of the system differs from input in that the data should be structured by the time maintenance is to be

performed. There are two methods: 1) line-by-line updating, and 2) line-within-paragraph updating.

LINE-BY-LINE (SIMP)

The SIMP (meaning simple) updating method provides for the addition, deletion, or changing of one or more lines of information at a time. A transaction to a line applies to all information contained therein. A file maintained by SIMP must be completely reproofed after each update, since the entire file is renumbered during the update.

LINE-WITHIN-PARAGRAPH (SIMPTEXT)

Like the SIMP method, SIMPTEXT alters one or more lines and applies to the alteration of the complete line. Unlike SIMP, however, the alterations take place within a specified paragraph, or unit of data. Thus it is not necessary to reproof the whole file after an update, but only altered units.

MANIPULATION

When data is deemed correct, we begin its manipulation. In some cases this is a minor function, but in others extensive manipulation is required. For example, if the data are to be composed in the same sequence as input, we need only transform it into a form acceptable to the composition system through our general data transformation program. If, on the other hand, indices are to be constructed or the data are not in a sequence amenable to composition, then the data manipulation commands and the variable input printer may be employed to achieve the desired structure and to provide proof listings of the results.

GENERALIZED DATA TRANSFORM

This program edits data in the maintenance formats and outputs flags and tags to identify hierarchy and typographical characteristics. The output of this program is usually input to typesetting or line printer proofs. Actually, this process merely converts mnemonics that are peculiar to a particular file into identifications explicit to the composing system.

GENERALIZED DATA INDEXER (INDEX)

The INDEX program enables outputting selected entries from a data unit, attached to which is a sorting field. It is assumed that sorting will take place after an INDEX run, and that the resultant file will be DEDUPED.

DATA MANIPULATION COMMANDS (DMC)

Rather than try and conform data to system requirements, we decided to create a programming method that can perform many tasks with a minimum of effort. The DMC represents successful implementation of this philosophy. As a programming language, DMC is able to bend to specific job requirements. As a high level text manipulation command language, it enables us to perform specific functions while expending a minimum of implementation time.

DMC programs are composed with essentially no regard to data length and core location characteristics. High level functions such as data stringing, table searches, word and phrase scanning, character-string translations, and Boolian and computational statements, provide an extremely flexible text manipulation capability.

VARIABLE INPUT PRINTER (VIP)

The VIP program is capable of producing formated printouts from any transformed file and can control column width, bold typesetting, and underscoring. This program is also employed to allow computer printouts to appear in a form essentially that of the composed data.

COMPOSITION

The RECOMP composition package provides a sophisticated output capability. Data that have been processed by the input, maintenance, and manipulation areas of the system can be typeset and printed on a variety of photocomposition devices, in page form and in a variety of type styles.

Generally, input comes from the transformation program, however, alternate sources, such as DMC, are available. In the RECOMP system, unlike others, data is not maintained with specific typographical information imbedded in the data file itself. The typesetting characteristics are input in the form of parameters outside the file, and typography is known to the typesetting system only. Thus, data maintenance is simplified and it is not necessary to create a separate composition file.

As previously mentioned, many devices may be chosen to print the composed file. Examples include the Photon 713 and 901, Linotron 1010, Videocomp 830 and APS-3.

There are four major components of the composition system: 1) definition, 2) galley makeup, 3) page makeup, and 4) device drivers.

DEFINITION

The definition program accepts key-word parameters that describe the format to be achieved. Such characteristics as typeface, point sizes, and leading are input at this time. As previously indicated, typesetting characteristics are not included in the file. The charac-

teristics defined at this point, therefore, refer to flags and tags that have been included during the transformation process.

GALLEY MAKEUP

The RECOMP galley makeup program reads transformed files and justifies them into galleys in accordance with the definitions provided via the definition program. All centering, justifying, point sizing, and leading characteristics are implemented by this program. The output is usually input to the pagination program, but may be directly input to one of the driver programs. It should be stressed that the output of galley makeup is device independent.

PAGE MAKEUP

The pagination program is capable of providing fully made-up pages from the output of galley makeup. Again, the parameters for pagination are input to the definition program. Optional page characteristics, such as running heads, number of and variation of the columns on a page, leading, and positioning of page numbers, are solidified by this program. The output of the page makeup program is still device independent, therefore, allowing utilization of a single pagination program for all devices.

DRIVERS

It is the function of the photocomp drivers to translate the output of the galley and page makeup programs into information that can be input to specific devices. For each photocomposition device we wish to operate, there is one driver program.

RETRIEVAL

Our retrieval system is a full-text system, i.e., the entire text is searched, not an abstract. The system operates on textual data bases that have been input and stored on direct access storage devices (DASD), disks, drum, etc. The system comprises three elements: 1) the data file, 2) the thesaurus, and 3) the computer programs.

Searches are performed and outputs are provided in a matter of seconds. System speeds are achieved through prudent thesaurus and file storage techniques. Outputs may be directed to console, tape, or high speed printer. The system may be operated in either an on-line or batched environment.

A search is initiated when a "Subset" statement is input on the console. Having subset the file, the user is informed of the number of elements that have qualified. At this time, he may choose to output the qualified data or further subset the file. Thus a two-cycle process of subsetting and outputting is continuously repeated. If

batched queries are utilized, it is, of course, unlikely that the user is present and, therefore, although output, the number of qualified elements is not of immediate importance. Upon examination of this information, however, future queries may be more rigidly structured.

FILES

Data files which are the sources of retrieval must be in the paragraph-by-paragraph maintenance format. Thus, a single body of text can be delimited and interrogated. (It should be noted that, although data are normally interrogated on a per-unit basis, it is possible to expand this to groups of units.)

Data elements are seldom unique and usually are only one of several related elements. For example, a bibliographic entry would be divided into author, title, collation and tracings. When retrieving units, it is usually desirable to output authors, titles and call numbers as higher level elements at a minimum, and possibly an entire bibliographic entry. To facilitate this, all elements of an entry are linked when stored on disk. This allows output from any part of the total document structure when a qualified unit is located.

All unit locations and linkages are maintained as identification data and never as actual device addresses. This allows the relocation of any unit or group of units with no effect on the retrieving system. Therefore, the maintenance of the data file may proceed in as simple a manner as possible, and with no real regard to its effect on the retrieval of information. The single connection between maintenance and retrieval is the file directory. This table equates all entry unit numbers to actual device locations and is employed by the retrieving programs when output is indicated.

THESAURUS

Autocomp has come in contact with a variety of data bases ranging from small directories to federal tax case citations to municipal and state codes; however, primarily it has dealt with law-oriented data bases. Experience has taught us that the most accurate system of retrieval is only as good as its thesaurus index to the data base. Thus, we have endeavored to provide machine-developed thesauri which are on one hand complete and on the other efficient and not overly bulky.

One method of constructing a thesaurus is to output phrases under the following definition: a phrase is a portion of text between and including two key words. Each phrase is output with the *number of the unit* from which it has been derived.

Undoubtedly, a large number of meaningless phrases are thus output. Therefore, prior to storing a thesaurus, it should be scanned and totally meaningless phrases should be eliminated or altered to

achieve relevance. At a minimum, this method produces an extremely accurate key-word index and is a simple means of extracting multiple-word key phrases. Another advantage to a thesaurus so generated is its total accuracy.

Whatever way the thesaurus of phrases is derived, it is stored on a disk in a sequence that facilitates rapid access to groups of phrases. Whenever units are added to a data file, they must be edited, and phrases resulting from the edit must be added to the thesaurus for that file.

It should be noted that it is the objective of the system to search the highly structured thesaurus which can produce qualified information on a timely basis, and not to search the text of the file itself.

PROGRAMS

Two extremely basic functions are performed by the retrieval programs: subsetting, and file outputting.

Subsetting is performed through the employment of a simple statement that directs the programs to locate and tabulate units containing a specific combination of phrases. For example, the statements "Tax" or "Taxes" and "Pay" direct the system to locate all entries pertaining to the payment of taxes. The statements "Vehicle" and "Licen" direct the system to locate entries pertaining to the licensing of vehicles. Of course, the higher level elements are available as well.

Outputting is also indicated via simple statements. For example, "List tracings" would indicate that titles and tracings only are to be output, while the statement "List entry" would cause the output to include an entire bibliographic notation that contains qualifying data. The user may specify "List" for high speed printer, "Type" for console, or "Save" for tape outputs.

Additionally, the user may specify "Hold" as the outputting function. In this case, the qualified units are not output, but are retained in a holding area for further qualification. The subset statement may then specify the phrase "Hold and" which will subsequently cause the holding units to be included as part of the criteria evaluation.

At Autocomp we have not developed a new system to handle MARC data, but instead are adapting a successful operational text-processing system to the peculiarities of MARC.

Frederick G. Kilgour
Director
The Ohio College Library Center
Columbus, Ohio

SUMMARY OF MARC

MARC is a label attached to an increasing variety of formats containing machine-readable cataloging information. However, MARC is first and foremost a format for interchange of bibliographic information, and BNB MARC and MARC II derive from the proposed "USA Standard for a Format for Bibliographic Information Interchange on Magnetic Tape."[1] If a format is to bear the label MARC it should have the capability of generating a standard communication format derived from the USA standard.

Although this standard in no way prescribes data contained within a MARC format, those data are unfortunately beginning to acquire the label. Hence, codes and cataloging information are also beginning to be known as MARC. Nevertheless, this paper will attempt to distinguish between format and content.

Applications of machine-readable cataloging information must surmount various categories of obstacles to achieve routine operation, and until the present time relatively little has been achieved if one views the library community as a whole. However, from the viewpoint of research and development the attainment is considerable and successful. Although the LC MARC project has been designed for the use of LC cataloging information, primarily for catalog card production, the data on MARC II tapes have been used not only for catalog card production, but also for bookform catalog production and a variety of other listings, such as acquisitions lists and selective dissemination of information (SDI) reports. It has also been used in the design of on-line catalogs.

Clearly, MARC II and BNB MARC are successes. They communicate bibliographic information in a useful format, and the bibliographic information itself is being used. Nevertheless, applications of the MARC format and its contents have revealed a few areas—surprisingly few—where evolutionary improvement may be expected. On the whole the MARC formats are astoundingly well designed.

It cannot be too often emphasized that the primary characteristic of MARC communication formats is physical; the USA standard prescribes only physical characteristics. The MARC communication format is analogous to the 75 by 125 mm. catalog card employed by the unit-card system with its standardized format for main entry, body of title, physical description, notes, and tracings. Like the

standardized unit card, the MARC communications format is designed primarily to facilitate interchange of cataloging information.[2]

PROCESSING PROBLEMS

The subcommittee that designed the USA standard attempted to produce a format for communications that would also be efficient for processing. On the basis of present experience, it appears that the committee almost, but not quite, attained this goal. The bringing together of directory information as the second logical group of data in the standard format has proved inefficient for space and processing. Relocation of tag and length of field information as a leader for each tagged element of data, thereby making possible omission of starting character position, greatly facilitates additions or subtractions to the record, for it is no longer necessary to recalculate affected starting character positions. This change is relatively minor, but it does reduce processing costs.

A beneficial, general principle for processing bibliographic data is that whenever possible, processing should be done only once and results that may be reused should be saved. At the present time, MARC II does not provide for storage of sort keys (whereas BNB MARC does provide some sort data) or for information useful in formating call numbers. The Library of Congress is currently working toward an amendment to the MARC II format that would provide for retention and communication of generated sort keys. Similarly it would be useful to provide a technique for communicating LC call numbers in the 050 field in a partially formated status. The Dewey classification number in the 082 field presents no problem.

A whole series of questions related to the minimum amount of data required for a useful MARC record still remain unresolved. For catalog construction these records may go into bookform catalogs, card catalogs, or on-line computerized catalogs. Each of these catalogs has different requirements. For use in a bookform catalog the sort key should be in sufficient detail to enable machine alphabetizing of all entries in one grand sort, but the cataloging data can be somewhat less complete than that required in a card catalog. For example, tracings serve no purpose in a bookform catalog, and there is somewhat less of a requirement for determining an exact position of a main entry since it is possible for the eye to glance over fifty to a hundred entries at a time.

A record to be used to produce a card to go into a card catalog requires a less precise sort key, since the cards will only be roughly alphabetized for filing. However, determination of the main entry must be more precise, for it is more important in a card catalog that the card be in its correct location. Also, it is necessary that at least one card in one catalog contains tracings that reveal the location of other cards describing the same book.

Although there has been little experience with on-line catalogs, it seems apparent that a sort key probably will not be used to arrange records in a file and that only the most rudimentary of sort keys would be required for possible alphabetizing of a score or more records for displaying on a cathode ray tube terminal. Since the user would almost always be working with small displayed catalogs containing only tens of items, it seems unnecessary to elaborate on bibliographic description beyond the originals on title pages and at the heads of articles.

From the previous paragraph it is clear that MARC records for on-line catalogs require considerably less data than do MARC records for production of catalog cards or bookform catalogs. Hence, the question arises as to how complete any given MARC record should be and still be usefully interchangeable. Clearly, MARC records for an on-line catalog can be more simple than those from which catalog cards or bookform catalogs would be produced. Therefore, it would burden those constructing on-line catalogs to require that their cataloging data also be adequate for manufacture of catalog cards and bookform catalogs. Rather, it appears that it would be more useful to define a minimum record that could be used as a base by another user to construct a more complicated record required for the second user's purposes.

Horseback opinions expressed in committee meetings will not be sufficient to determine such a minimum standard. Rather, it will be necessary to design in detail a MARC system based on minimal data. Once such a system has been designed and tested, the minimum standard can be established.

Although MARC II provides for abridged titles in accordance with Rule 133B of the *Anglo-American Cataloging Rules*,[3] it does not provide for production of abridged titles as added entries when the abridgement does not include the first words of the title. A significant portion of libraries employ abridged titles or striking titles that do not include first words of the title, thus the addition of this capability to the MARC II system would enhance it for many users.

PROBLEMS WITH CALL NUMBERS

Cataloging codes in the twentieth century have largely neglected subject added entries and appear to have entirely neglected call numbers. This absence of rules for the construction of call numbers, associated location stamps, and holdings information generates the most difficult problems to solve in the computerized production of catalog cards. In the OCLC's (Ohio College Library Center) catalog production system it has been necessary to add subfield codes only in call number fields. OCLC employs "$*" to indicate that a call number has been partially formated, "$+" to identify a location

stamp, "$-" to identify holdings information, and a "$," to indicate that a class number in a call number field is not followed by a book number.

Because of the seemingly uncontrolled variety of information included in Library of Congress call numbers, it can be said that LC call numbers are pseudo free text. Therefore, it is not possible to parse LC call numbers with confidence. Three examples of bizarre call numbers follow:

Card Number	Call Number
64-62399	$aJ82$bD792 May 8, 1964
74-604824	$aQE75$b.P9 no. 280 etc.
70-449376	$aJX197$b.A2 E/ECE/665

If call number data are to approach free text as LC call numbers now do, it will be most helpful in the mechanized formating of such call numbers if a separator code is placed at the end of each logical segment of the number.

Call number problems in MARC II data are further enhanced by the Library of Congress's practice of placing in the 050 field—Library of Congress call number—whatever the Library of Congress uses as a call number. It is suggested that the term Library of Congress call number should be defined as a Library of Congress classification number to which a derived book number has been added thereby yielding a call number. It is this type of call number that is useful to other libraries.

However, MARC II records emanating from the Library of Congress contain three other types of information in this field. First, there is the case of a Library of Congress call number as just defined, but followed by a $a and an alternate class number. This condition occurs most frequently, but not necessarily entirely, in the case of LC PZ3 and PZ4 call numbers. Second, there is the case of only a classification number appearing, and third, the word "Law" is a frequent occupant of the 050 field.

None of these three types of data are Library of Congress call numbers as ordinarily defined, and none can be used as they exist in the Library of Congress call number field as call numbers on cards for other libraries. Therefore, in processing the 050 field it is necessary to determine if this LC call number field actually contains an LC call number, or what part of the data is an LC call number. The OCLC catalog production system puts out a unit card whenever case two or three occurs.

On the MARC II weekly tape numbered V2N002 that was distributed on April 2, 1970, 12 percent of the records bore either case one, two, or three types of data. However, if it were only 1 percent of the

records that had non-Library of Congress call numbers in the 050 field, it would still be necessary to expend needless processing time on every record to determine the character of the data in the 050 field. It is suggested that the MARC II record would be improved by designating a new separate field for each of the three cases presented in the previous paragraph, thereby enhancing the efficiency of processing.

CATALOGING DATA

There are no apparent problems associated with the computer processing of MARC II cataloging data. To be sure, the error rate in cataloging data is uncomfortably high, and it has been rumored that an analogous circumstance exists in BNB MARC records. Presumably the present frequency of errors in both LC MARC and BNB MARC will diminish when the two institutions begin routine applications of their records for their own purposes.

MARC II contains Library of Congress cataloging, and as Richard E. Coward has pointed out,[4] this cataloging does not always conform to Anglo-American cataloging rules because of the inability of the Library of Congress for administrative reasons to adopt these rules in their entirety. Therefore, MARC II cataloging data do not conform to either American or international cataloging standards.

The existence of nonstandard cataloging data in MARC II raises a problem that should be resolved. It is inconceivable that the Anglo-American rules can be discarded at this time. However, it is not inconceivable that the Library of Congress could prepare cataloging data for MARC II that would conform with the Anglo-American rules, or at the very least could indicate in a record a cataloging element or elements not in conformity with the rules. Resolution of this problem would enhance the international acceptance of a MARC format, as Coward has already shown.

Since BNB MARC and MARC II were created, experience makes it abundantly clear that the MARC-type format for machine-readable cataloging information is a success. In the case of MARC II, there appear to be only three relatively small problems associated with computer processing. First, there is the awkwardness of the directory being located in front of all the data instead of being spread out through the data with each pertinent piece of directory information being located at the beginning of each field. Second, there is the problem of the generation and retention of sort keys on which the Library of Congress is now working. Third, there is the problem of formating call numbers. This problem exists because of the lack of detailed rules for setting up call numbers, and it is suggested that such rules be drawn up.

MARC II cataloging information does not conform with the

Anglo-American rules. This nonconformity can generate problems for libraries that have established the policy of conforming to Anglo-American rules in their entirety. Hence the problem should be resolved, and there are presumably several avenues that could lead to an effective resolution.

The MARC format is a standardized format and like all standards can be expected to evolve. It is gratifying that at this early moment in the history of MARC there appears to be minimal evolutionary pressure to attain full success in a classical cataloging environment.

REFERENCES

1. United States of America Standards Institute. Z39 Committee. Subcommittee on Standardization in the Field of Library Work and Documentation. "USA Standard for a Format for Bibliographic Information Interchange on Magnetic Tape," *Journal of Library Automation*, 2:53-65, June 1969.

2. Kilgour, Frederick G. "Standardization for Interchange of Cataloging Records—MARC II. In *Proceedings of the Third International Congress on Medical Librarianship.* (In press.)

3. *Anglo-American Cataloging Rules.* Prepared by the American Library Association, the Library of Congress, the Library Association, and the Canadian Library Association. North American Text. Chicago, ALA, 1967, p. 195.

4. Coward, Richard E. "MARC International," *Journal of Library Automation*, 2:181-86, Dec. 1969.

INDEX

A

Argonne National Laboratory, distribution of MARC tapes, 27-31.
Autocomp, Inc., adaptation of RECOMP to use MARC, 98-104.
Automation of libraries, difficulties of, 24-25; effect of in Great Britain, 44, 46, 47, 48-49; importance of in handling files, 2.
Avram, Henriette D., "The Evolving MARC System: The Concept of a Data Utility," 1-26.

B

Becker, Joseph, 57.
Bibliographic system in Washington State Library, 61; subsystems to, 62-63.
"The BNB MARC Project," R. E. Coward, 36-52.
Book catalogs, difficulties in producing, 106-107; distribution of, 63; production of, v, 3, 56, 59-61, 62.
Book trade, automation of in Great Britain, 39-40.
British MARC Project, development of subject index, 43-44; difficulties in designing national bibliography, 40-43; difficulties in implementing, 48-49; difficulties of standardization, 49-51; effects of, 44, 46; importance of BNB in, 37-39; national versus international level, 36-37, 42-43, 49-50; role in book trade, 39-40; types of users, 42-43.
British National Bibliography (BNB), 36, 37; problems of implementing MARC, 40-43, 48-49; production of trade bibliography, 39-40; use of PCMI in production of, 47-48; utilization of MARC, 38-39.
Bucket code, 68.

C

Cole, Fred C., Jr., "Commercial Users of MARC: Autocomp, Inc.," 98-104.
"Commerical Users of MARC: Autocomp, Inc.," Fred C. Cole, Jr., 98-104.
Computer output microfilming, v-vi.
Consonant-string code, 67.
Content designators in MARC, 4, 9-11, 15, 16-17.
Control codes, in British MARC, 42; in MARC, 9-11, 15, 16-17.
Coward, R. E., "The BNB MARC Project," 36-52.
Cronin, John, 36.

D

Data utility, advantages of, 3; applications of MARC system to, 3, 9-24; concept of, 1-3; handling of files in, 2-3; versus computer utility, 2.

E

"The Evolving MARC System: The Concept of a Data Utility," Henriette D. Avram, 1-26.

F

Files, accessing of, 19, 21-24, 70-73; organization of, 19, 20; updating of, 5, 14, 17-18, 81, 100.
Format recognition process, 15-17.

G

Griffin, Hillis, "MARC Users: A Study of the Distribution of MARC Tapes and the Subscribers to MARC," 27-35.

H

Hashed files, 70-72.
Hayes, Robert, 57.
Henderson, Kathryn Luther, "Introduction," ix-x.

I

IBM Index Sequential Accessing Method, 19.
Indexed register system at Rice University, 73-75.
Indexing systems, of BNB MARC Project, 43-44; of CUMARC, 83, 85, 87, 92; of MARC, 20-24; of Washington State Library network, 59-61, 62.
Interlibrary cooperation in Washington State, 54-56.
International MARC, conflicts of users, 42; difficulties in creating, 36-37, 109; standardization of, 49-52.
International Standard Book Number, 39.

K

Kilgour, Frederick G., "Summary of MARC," 105-110.

L

Leonard, Lawrence, 31.
Library Services and Construction Act, 54, 55, 57.

M

McGee, Robert S., and Miller, Robert C., "MARC Utilization in the University of Chicago Library Bibliographic Data Processing System," 78-97.
"MARC at Rice University," Frederick H. Ruecking, Jr., 66-77.
"MARC Book Catalog Production in Washington State," Josephine S. Pulsifer, 53-65.
MARC Distribution Service, 1, 8, 9, 20, 58, 92-97; role of Argonne National Laboratory in, 27-31; programs used in, 29-30; use of tapes, 27-29.
MARC programs, descriptions of, 4-8, 15-19, 100-102, 104.
MARC subject descriptor string, 44.
MARC system, description of, 3-8, 105; difficulties of, 8, 92-95, 106-109; expansion of, 9-24, 96-97, 107, 109; problems of international standardization, 36-37, 42-43, 49-51, 109-110; rationale of, 1-3; utilization of at Rice University, 66-77; utilization of at the University of Chicago Library, 78-92; utilization of at the Washington State Library, 59-61, 63-64; utilization of in data utility, 3-9.
MARC tapes, 27; cleaning of, 28; copying of, 30; cost of, 28-29; mailing of, 30; problems of, 28; reuse of, 29.
"MARC Users: A Study of the Distribution of MARC Tapes and the Subscribers to MARC," Hillis Griffin, 27-35.
MARC users, problems of, 31-35; solutions for, 33-35; types of, 42.
"MARC Utilization in the University of Chicago Library Bibliographic Data Processing System," Robert S. McGee and Robert C. Miller, 78-97.
Master file, accessing of, 21-24; organization of, 20.
Micrographic storage, 76-77.
Milestone Tables, 19-20.
Miller, Robert C., and McGee, Robert S., "MARC Utilization in the University of Chicago Library Bibliographic Data Processing System," 78-97.
Multiple-ring file, 73.

P

Pacific Northwest Bibliographic Center, description of, 54-55; sharing of resources, 55-56; sharing of technical sources, 56.
Phonemic code, 68.
Photo-Chromic Micro Image (PCMI), use of in Great Britain, 47-48.
Plato IV, vi.
Pulsifer, Josephine S., "MARC Book Catalog Production in Washington State," 53-65.

R

RECOMP, composition of data, 101-102; input of data, 98-101; maintenance of data, 99-100; manipulation of data, 100-101; processing of data, 98-104; retrieval of data, 102-104.
RECON Pilot Project, 11, 58-59.
Rice University, access technique studies, 70-73; organization of indexed register system, 73-77; search code studies, 66-70; use of MARC, 66-77.
Ruecking, Frederick H., Jr., "MARC at Rice University," 66-77.

S

Search codes, 66-67, 68-70; types of, 67-68.
Spaulding, Carl M., "Foreword: The Meaning of MARC," vii-viii.
Standard Book Numbering (SBN) system, as a MARC data element, 96; in Great Britain, 39; on the international level, 39.
Structural tree filing of data, 72-73, 74.
"Summary of MARC," Frederick G. Kilgour, 105-110.

T

Three-to-one code, 67-68.

U

Universal Decimal Classification (UDC), 46, 51-52.
University of Chicago Library system, planned expansion of, 89-92; problems of, 89-91; statistics on initial usage of MARC, 83-89; utilization of MARC, 78-83.
Updating of files, 5, 14, 17-18, 81, 100.

W

Washington Higher Education Library Committee, 55.
Washington Library Association, 54, 55.
Washington library network, as proposed by Becker and Hayes plan, 57-58; development of bibliographic system, 61-63; implementation of MARC, 59-61, 64; origin of concept, 56-57; requirements for, 58-59; technical aspects of, 64.
Washington State, development of library services, 54-56; profile of libraries, 53.
Washington State Library Commission, 54, 55, 57.
Working file, accessing of, 19; organization of, 19.

UNIVERSITY OF ILLINOIS-URBANA
3 0112 044715867